

**MASTER COPY:** PLEASE KEEP THIS "MEMORANDUM OF TRANSMITTAL" BLANK FOR REPRODUCTION PURPOSES. WHEN REPORTS ARE GENERATED UNDER THE ARO SPONSORSHIP, FORWARD A COMPLETED COPY OF THIS FORM WITH EACH REPORT SHIPMENT TO THE ARO. THIS WILL ASSURE PROPER IDENTIFICATION. NOT TO BE USED FOR INTERIM PROGRESS REPORTS; SEE PAGE 2 FOR INTERIM PROGRESS REPORT INSTRUCTIONS.

**MEMORANDUM OF TRANSMITTAL**

U.S. Army Research Office  
ATTN: AMSRL-RO-BI (TR)  
P.O. Box 12211  
Research Triangle Park, NC 27709-2211

☐ Reprint (Orig + 2 copies)

☐ Technical Report (Orig + 2 copies)

☒ Manuscript (1 copy)

☐ Final Progress Report (Orig + 2 copies)

☐ Related Materials, Abstracts, Theses (1 copy)

CONTRACT/GRANT NUMBER: W911NF0410224 (46637CIMUR)

REPORT TITLE: Power Allocation for a MIMO Relay System  
with Multiple-Antenna Users

is forwarded for your information.

Accepted in:

IEEE Transactions on Signal Processing

Sincerely,

Dr. James Zeidler  
Department of Electrical and Computer Engineering  
University of California, San Diego

<b>REPORT DOCUMENTATION PAGE</b>			Form Approved OMB NO. 0704-0188	
Public Reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comment regarding this burden estimates or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188,) Washington, DC 20503.				
1. AGENCY USE ONLY ( Leave Blank)		2. REPORT DATE Jan. 2010		3. REPORT TYPE AND DATES COVERED Manuscript 2010
4. TITLE AND SUBTITLE Power Allocation for a MIMO Relay System with Multiple-Antenna Users			5. FUNDING NUMBERS W911NF0410224 (46637CIMUR)	
6. AUTHOR(S) Yuan Yu and Yingbo Hua				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) University of California – San Diego Department of Electrical and Computer Engineering 9500 Gilman Dr., La Jolla, CA 92093			8. PERFORMING ORGANIZATION REPORT NUMBER N/A	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U. S. Army Research Office P.O. Box 12211 Research Triangle Park, NC 27709-2211			10. SPONSORING / MONITORING AGENCY REPORT NUMBER N/A	
11. SUPPLEMENTARY NOTES The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.				
12 a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution unlimited			12 b. DISTRIBUTION CODE N/A	
13. ABSTRACT (Maximum 200 words)  A power allocation or scheduling problem 1 is studied for a multiuser MIMO wireless relay system where there is a non-regenerative relay between one access point and multiple users. Each node in the system is equipped with multiple antennas. The purpose of this study is to develop fast algorithms to compute the source covariance matrix (or matrices) and the relay transformation matrix to optimize a system performance. We consider the minimization of power consumption subject to rate constraint and also the maximization of system throughput subject to power constraint. These problems are nonconvex and apparently have no simple solutions. In this paper, a number of computational strategies are presented and their performances are investigated. Both uplink and downlink cases are considered. The use of multiple carriers is also discussed. Moreover, a generalized water-filling (GWF) algorithm is developed to solve a special class of convex optimization problems. The GWF algorithm is used for two of the strategies shown in this paper.				
14. SUBJECT TERMS Network of MIMO links, medium access control, space-time power allocation, space-time power scheduling, multiuser MIMO relays, convex optimization, non-convex optimization, generalized water filling.			15. NUMBER OF PAGES 29	
			16. PRICE CODE N/A	
17. SECURITY CLASSIFICATION OR REPORT <b>UNCLASSIFIED</b>	18. SECURITY CLASSIFICATION ON THIS PAGE <b>UNCLASSIFIED</b>	19. SECURITY CLASSIFICATION OF ABSTRACT <b>UNCLASSIFIED</b>	20. LIMITATION OF ABSTRACT  <b>U</b>	

NSN 7540-01-280-5500

Standard Form 298 (Rev.2-89)  
Prescribed by ANSI Std. Z39-18  
298-102

Enclosure 1

# Power Allocation for a MIMO Relay System with Multiple-Antenna Users

Yuan Yu and Yingbo Hua, *Fellow, IEEE*

## Abstract

A power allocation or scheduling problem<sup>1</sup> is studied for a multiuser MIMO wireless relay system where there is a non-regenerative relay between one access point and multiple users. Each node in the system is equipped with multiple antennas. The purpose of this study is to develop fast algorithms to compute the source covariance matrix (or matrices) and the relay transformation matrix to optimize a system performance. We consider the minimization of power consumption subject to rate constraint and also the maximization of system throughput subject to power constraint. These problems are non-convex and apparently have no simple solutions. In this paper, a number of computational strategies are presented and their performances are investigated. Both uplink and downlink cases are considered. The use of multiple carriers is also discussed. Moreover, a generalized water-filling (GWF) algorithm is developed to solve a special class of convex optimization problems. The GWF algorithm is used for two of the strategies shown in this paper.

## Index Terms

Network of MIMO links, medium access control, space-time power allocation, space-time power scheduling, multiuser MIMO relays, convex optimization, non-convex optimization, generalized water filling.

Copyright (c) 2008 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to [pubs-permissions@ieee.org](mailto:pubs-permissions@ieee.org).

Y. Yu and Y. Hua (corresponding author) are with the Department of Electrical Engineering, University of California, Riverside, CA, 92521. Emails: [yyu@ee.ucr.edu](mailto:yyu@ee.ucr.edu) and [yhua@ee.ucr.edu](mailto:yhua@ee.ucr.edu). This work was supported in part by the U. S. Army Research Office under the MURI Grant No. W911NF-04-1-0224, the U. S. Army Research Laboratory under the Collaborative Technology Alliance Program, and the U. S. National Science Foundation under Grant No. TF-0514736.

<sup>1</sup>The two terms “power scheduling” and “power allocation” are used interchangeably in many cases in the literature. But the former stresses the computation of the latter in advance.

## I. INTRODUCTION

Wireless relays are known to be useful to increase the coverage of wireless communications under power and spectral constraints. A wireless relay can be regenerative or non-regenerative. A regenerative relay requires digital decoding and re-encoding at the relay, which can cause a significant increase of delay and complexity. A non-regenerative relay does not need any digital decoding and re-encoding at the relay, which is a useful advantage over regenerative relays.

Recently, there have been many research efforts on non-regenerative MIMO relay systems [1], [2], [3], [4], [5], [6], [7], [8], [9]. A non-regenerative MIMO relay applies a transformation matrix, also called relay matrix, to its received signal vector and then forwards it to the next node. The MIMO relay formulation in [3] includes the multicarrier relay problem in [10] as a special case. This paper continues to address non-regenerative MIMO relay systems. In particular, we consider power allocation problems. In the context of MIMO relays, a power allocation problem is about the determination of the source covariance matrix and the relay matrix to maximize a system performance.

For a single-user two-hop MIMO relay system, an optimal structure of the relay matrix that maximizes the source-to-destination mutual information was presented in [1] and [2], and an optimal structure for both the source covariance matrix and the relay matrix was established in [3]. The optimality of this structure, which is essentially a diagonalization or decoupling of the entire relay system into a set of parallel scalar sub-systems, is recently established in [5] for a broader class of objective functions known as Schur-convex or Schur-concave functions. Furthermore, this elegant structure is also shown in [6] to be optimal for a multi-hop MIMO relay system of any number of hops.

For multiuser MIMO relay systems, however, the above mentioned property does not hold any more. Finding the source covariance matrix and the relay matrix to maximize a system performance is generally a difficult task. Prior efforts on multiuser MIMO relay systems are reported in [7], [8] and [9]. In these works, each user is assumed to have a single antenna. Part of the reason for this assumption was to simplify the problem. Additional references on MIMO relays can be found in [11].

In this paper, we focus on a multiuser two-hop MIMO relay system where each node is equipped with multiple antennas. For this problem, not only the diagonal structure as shown in [1], [2], [3] and [5] is no longer optimal, but also the uplink-downlink duality property shown in [12] and [9] no longer applies. This makes the optimal power allocation a difficult task. Facing the challenge unsolved by others, we will present a number of computational strategies to search for the best possible power allocation. We will consider both uplink and downlink problems. We will also consider both system throughput

maximization and power consumption minimization. These algorithms are summarized in Table I and discussed in detail in this paper. These algorithms are designed to solve the power allocation problems more general than those treated before. In particular, for a problem treated in [7], our approach can yield much better results than the approach developed there.

We assume that all channel matrices are known to a central scheduler and to the transmitters and receivers if needed. Except for Algorithm 1, all other algorithms in Table I are not mathematically proven to yield globally optimal results for their corresponding problems. However, Algorithm 1 is based on a reformulation of the original problem, which essentially approximates the original non-convex problem by a convex problem. Because of this approximation, there is a significant penalty to the performance of Algorithm 1 as shown later in Section VI.

We will also develop a generalized water filling (GWF) theorem and the corresponding GWF algorithm to solve with global optimality a special type of convex optimization problems. The GWF algorithm is a useful building block for two of the power allocation algorithms summarized in Table I. In the literature there are other types of algorithms also called generalized water filling. But they were actually designed for different problems. Our GWF algorithm is a generalization of the conventional water filling algorithm from single power constraint to multiple power constraints.

In Section II, the GWF theorem is presented. In Section III, we treat a multiuser MIMO relay downlink system. We present power allocation algorithms for maximizing the system throughput (i.e., sum rate) under a power constraint, and power allocation algorithms for minimizing the system power consumption under individual user rate constraints. In Section IV, we deal with similar issues for the uplink case. In Section V, we show how to apply our algorithms for joint multicarrier power allocation. In Section VI, simulations results are presented to illustrate the performances of our algorithms. This study confirms that power allocation affects the system performance significantly and developing fast algorithms for power allocation is critically important.

## II. A GENERALIZED WATER-FILLING ALGORITHM

Consider the following convex optimization problem:

$$\begin{aligned} \min_{\mathbf{Q} \succeq 0} \quad & J \doteq -\log |\mathbf{I} + \mathbf{H}\mathbf{Q}\mathbf{H}^H| \\ \text{s.t.} \quad & \text{tr}\{\mathbf{B}_i\mathbf{Q}\mathbf{B}_i^H\} \leq P_i, \quad \forall i \in \{1 \dots m\} \end{aligned} \quad (1)$$

where  $\mathbf{H}$  and  $\mathbf{B}_i$  are complex matrices,  $\mathbf{Q}$  is a complex positive semi-definite matrix, and  $P_i$  are positive numbers. Without its base specified,  $\log$  has the natural base  $e$ . If  $m = 1$ , the solution to the above

problem can be found by a well known water-filling algorithm. It is a fast algorithm for this particular case. If  $m > 1$ , however, there appears no fast algorithm available in the prior literature except for the general purpose convex optimization programs such as the CVX package designed for Matlab [13]. We now introduce a special purpose algorithm, referred to as generalized water-filling (GWF) algorithm, to solve the problem in (1). The GWF algorithm is based on the following GWF theorem:

*Theorem 1:* The solution to (1) is given by:

$$\mathbf{Q} = \mathbf{K}^{-H} \mathbf{V} (\mathbf{I} - \mathbf{\Sigma}^{-2})^+ \mathbf{V}^H \mathbf{K}^{-1} \quad (2)$$

where  $\mathbf{K} = (\sum_{i=1}^m \mu_i \mathbf{B}_i^H \mathbf{B}_i)^{1/2}$  (assumed to be non-singular),  $\mathbf{V}$  and  $\mathbf{\Sigma}$  are determined from the SVD  $\mathbf{H}\mathbf{K}^{-H} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$ ,  $(\cdot)^+$  replaces all negative diagonal elements by zeros and leaves all non-negative diagonal elements unchanged, and  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_m)$  are the solution to the following dual problem:

$$\begin{aligned} \max_{\boldsymbol{\mu} \geq 0} \quad & -\log |\mathbf{I} + \mathbf{H}\mathbf{Q}\mathbf{H}^H| + \sum_{i=1}^m \mu_i (\text{tr}(\mathbf{B}_i \mathbf{Q} \mathbf{B}_i^H) - P_i) \\ \text{s.t.} \quad & \mathbf{Q} = \mathbf{K}^{-H} \mathbf{V} (\mathbf{I} - \mathbf{\Sigma}^{-2})^+ \mathbf{V} \mathbf{K}^{-1}. \end{aligned} \quad (3)$$

To our knowledge, this theorem is new. The proof of this theorem and an algorithm for computing  $\boldsymbol{\mu}$  are given in Appendices A and B, respectively. A complete Matlab script of the GWF algorithm is available at <http://www.ee.ucr.edu/~yhua/GWF.pdf>. As illustrated by a simulation example in Appendix C, the GWF algorithm can achieve the same accuracy as CVX, and the former has a much faster speed than the latter when the dimension of  $\boldsymbol{\mu}$  is much smaller than that of  $\mathbf{Q}$ . The GWF algorithm is useful for more applications than those shown in this paper. For example, if one wants to design a source covariance matrix to maximize the data rate of a MIMO link and also wants to keep the interference from this source to other neighboring nodes under certain limits, such a problem can be directly formulated as (1).

### III. MULTIUSER MIMO DOWNLINK RELAY

We first consider the multiuser MIMO downlink relay system as illustrated in Fig. 1, where  $\mathbf{x} \in \mathcal{C}^{M \times 1}$  denotes the signal transmitted from the source equipped with  $M$  antennas,  $\mathbf{F} \in \mathcal{C}^{M \times M}$  the transformation matrix performed by the non-regenerative relay also equipped with  $M$  antennas, and  $\mathbf{y}_i \in \mathcal{C}^{N \times 1}$  the signal received by the user  $i$  equipped with  $N$  antennas. Furthermore,  $\mathbf{H} \in \mathcal{C}^{M \times M}$  denotes the channel matrix between the source and the relay,  $\mathbf{H}_i \in \mathcal{C}^{N \times M}$  is the channel matrix between the relay and the user  $i$ , and  $\mathbf{n}, \mathbf{n}_1, \dots, \mathbf{n}_K$  are the zero-mean Gaussian noises at the relay and the  $K$  users. Here, we assume that all the users are equipped with the same number of antennas. The transmission from the source to the relay is assumed to be orthogonal (in time and/or frequency) to the transmission from the relay to

all users. We also assume that the direct link between the source and any of the users is very weak and negligible.

Note that if the actual numbers of antennas at the users, relay or source are different from what is described above, we can always add imaginary dummy antennas to make up the number  $M$  or  $N$ . The effective  $\mathbf{H} \in \mathcal{C}^{M \times M}$  or  $\mathbf{H}_i \in \mathcal{C}^{N \times M}$  may have zero rows or zero columns, which however do not affect the expressions of our results.

The signal  $\mathbf{y}$  received at the relay, the signal  $\mathbf{r}$  transmitted from the relay, and the signal  $\mathbf{y}_i$  received by the user  $i$  can be expressed as follows:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \quad (4)$$

$$\mathbf{r} = \mathbf{F}\mathbf{y} = \mathbf{F}\mathbf{H}\mathbf{x} + \mathbf{F}\mathbf{n} \quad (5)$$

$$\mathbf{y}_i = \mathbf{H}_i\mathbf{r} + \mathbf{n}_i = \mathbf{H}_i\mathbf{F}\mathbf{H}\mathbf{x} + \mathbf{H}_i\mathbf{F}\mathbf{n} + \mathbf{n}_i \quad (6)$$

If  $\mathbf{n}$  has a covariance matrix  $\mathbf{C}_n$ , we can write  $\mathbf{C}_n^{-1/2}\mathbf{y} = \mathbf{C}_n^{-1/2}\mathbf{H}\mathbf{x} + \mathbf{C}_n^{-1/2}\mathbf{n}$  where the noise term  $\mathbf{C}_n^{-1/2}\mathbf{n}$  has the covariance matrix equal to the identity matrix. So, provided that the noise covariance matrices of  $\mathbf{n}$  and  $\mathbf{n}_i$  are known, we can assume for convenience that they are the identity matrices. We now define  $\mathbf{H}_c^H = [\mathbf{H}_1^H, \dots, \mathbf{H}_K^H]$ ,  $\mathbf{y}_c^H = [\mathbf{y}_1^H, \dots, \mathbf{y}_K^H]$  and  $\mathbf{n}_c^H = [\mathbf{n}_1^H, \dots, \mathbf{n}_K^H]$ . Then, using (6) for all  $i$ , we have

$$\mathbf{y}_c = \mathbf{H}_c\mathbf{F}\mathbf{H}\mathbf{x} + \mathbf{H}_c\mathbf{F}\mathbf{n} + \mathbf{n}_c \quad (7)$$

This is an effective channel model between the source and all users.

#### A. Maximization of Sum Rate under Power Constraint and ZFDPC (Algorithms 1-2)

The problem of maximizing the sum rate for all users under a power constraint for the downlink case was considered in [7] where each user has a single antenna. The authors also assume the use of zero forcing dirty paper coding (ZFDPC) [14]. We now extend the approach in [7] to users with multiple antennas.

Define the QR decomposition of the  $KN \times M$  matrix  $\mathbf{H}_c$  as  $\mathbf{H}_c = \mathbf{R}\mathbf{Q}$ , where  $\mathbf{Q}$  is an  $M \times M$  unitary matrix (which is not the same  $\mathbf{Q}$  in section II) and  $\mathbf{R}$  is a  $KN \times M$  lower triangular matrix. Define the SVD of the channel matrix  $\mathbf{H}$  as  $\mathbf{H} = \mathbf{U}_h\mathbf{\Sigma}_h\mathbf{V}_h^H$  where  $\mathbf{\Sigma}_h = \mathbf{\Lambda}_h^{1/2} = \text{diag}(\lambda_{h,1}, \lambda_{h,2}, \dots, \lambda_{h,N})^{1/2}$  with descending diagonal elements, and  $\mathbf{U}_h$  and  $\mathbf{V}_h$  are unitary.

We assume that the source precoder generates  $\mathbf{x} = \mathbf{A}_x\mathbf{s}$  where  $\mathbf{s}$  contains i.i.d. symbols of unit variance and  $\mathbf{A}_x$  is such that the source covariance matrix is  $\mathbf{\Pi}_x = E\{\mathbf{x}\mathbf{x}^H\} = \mathbf{A}_x\mathbf{A}_x^H = \mathbf{V}_h\mathbf{\Lambda}_x\mathbf{V}_h^H$

with  $\mathbf{\Lambda}_x = \text{diag}(\lambda_{x,1}, \lambda_{x,2}, \dots, \lambda_{x,M})$ . We also assume that the relay matrix is constructed as

$$\mathbf{F} = \mathbf{Q}^H \mathbf{\Sigma}_f \mathbf{U}_h^H, \quad \mathbf{\Sigma}_f = \mathbf{\Lambda}_f^{1/2} = \text{diag}(\lambda_{f,1}, \lambda_{f,2}, \dots, \lambda_{f,N})^{1/2} \quad (8)$$

Here, the source covariance matrix is matched to the right singular vectors of the channel matrix  $\mathbf{H}$ , the optimality of which for a single user relay system is shown in [3]. The relay matrix here is matched to the left singular vectors of  $\mathbf{H}$  and the unitary matrix  $\mathbf{Q}$  of  $\mathbf{H}_c$ , which is adopted only heuristically without proof of optimality. As mentioned in [7], the matrix  $\mathbf{Q}$  is also affected by column permutations of  $\mathbf{H}_c$ , which can be further optimized. With the above structures of the precoder  $\mathbf{A}_x$  and the relay matrix  $\mathbf{F}$ , (7) becomes

$$\mathbf{y}_c = \mathbf{R} \mathbf{\Sigma}_f \mathbf{\Sigma}_h \mathbf{s} + (\mathbf{R} \mathbf{\Sigma}_f \tilde{\mathbf{n}} + \mathbf{n}_c) \quad (9)$$

where  $\tilde{\mathbf{n}} = \mathbf{U}_h^H \mathbf{n}$ . Note that each element of  $\mathbf{s}$  represents a scalar stream of data. Since  $\mathbf{R}$  is lower triangular, it is clear from the first term of (9) that the interference from stream  $j$  to stream  $i$  for  $j > i$  is now absent. To remove the interference from stream  $j$  to stream  $i$  for  $j < i$ , we can use the dirty paper coding (DPC) starting from the first stream that corresponds to the first element of  $\mathbf{s}$  in (9). For the first stream, there is no interference from other streams and the conventional coding is applied. For the second stream, there is the interference from the first stream which is however known to the encoder. With DPC, the interference from the first stream to the second stream can be virtually eliminated. The same principle applies to the remaining streams. Then, with DPC, the effective signal to noise ratio for the  $i$ th data stream is

$$SNR_i = \frac{|R_{i,i}|^2 \lambda_{f,i} \lambda_{h,i} \lambda_{x,i}}{\sum_{j=1}^i |R_{i,j}|^2 \lambda_{f,j} + 1} \quad (10)$$

where  $R_{i,j}$  is the  $(i, j)$ th element of  $\mathbf{R}$ . Note that the use of DPC has removed the mutual interference between the elements of  $\mathbf{s}$ . But the first term (the sum) in the denominator of (10) is due to the noise forwarded from the relay. The above interference cancellation method based on the QR decomposition and the DPC is known as zero forcing dirty paper coding (ZFDPC) [14].

The problem of maximizing the sum rate of this downlink relay system under ZFDPC can now be formulated as

$$\max_{\mathbf{\Lambda}_f, \mathbf{\Lambda}_x} R'_{sum,d} \doteq \sum_i^{KN} \log_2(1 + SNR_i) \quad (11)$$

$$s.t. \quad \text{tr}\{\mathbf{\Lambda}_x\} \leq P_x \quad (12)$$

$$\text{tr}\{\mathbf{\Lambda}_f(\mathbf{\Lambda}_h \mathbf{\Lambda}_x + \mathbf{I})\} \leq P_f \quad (13)$$

where the power constraint (12) is for the source, and the power constraint (13) is for the relay. In [7], the problem (11) is solved by a geometric programming under a high SNR approximation, which will be referred to as Algorithm 1. Note that a weighted sum rate can be used for all sum rate maximization algorithms. But for convenience, we choose the unit weights.

Next, we present an algorithm without the high-SNR assumption, referred to as Algorithm 2. We will search for  $\mathbf{\Lambda}_f$  and  $\mathbf{\Lambda}_x$  in an alternate fashion, where each cycle of the alternation is as follows.

1) *Source optimization with fixed  $\mathbf{\Lambda}_f$* : It is easy to verify that with any fixed  $\mathbf{\Lambda}_f$ , the problem (11) is a special case of the problem (1) shown in Section II, and hence the optimal  $\mathbf{\Lambda}_x$  can be found by the GWF algorithm.

2) *Relay optimization with fixed  $\mathbf{\Lambda}_x$* : With any fixed  $\mathbf{\Lambda}_x$ , the optimal  $\mathbf{\Lambda}_f$  can be found by maximizing the following penalized function of (11):

$$L_1(\mathbf{\Lambda}_f) \doteq \sum_i^{KN} \log_2 \left( 1 + \frac{|R_{i,i}|^2 \lambda_{f,i} \lambda_{h,i} \lambda_{x,i}}{\sum_{j=1}^i |R_{i,j}|^2 \lambda_{f,j} + 1} \right) + \frac{1}{t} \left[ \log \left( P_f - \sum_i \lambda_{f,i} (\lambda_{h,i} \lambda_{x,i} + 1) \right) \right] \quad (14)$$

where the second term is the logarithmic barrier function [15] associated with the constraint (13). For convenience, we will also write  $L_1(\mathbf{\Lambda}_f) = L_1(\boldsymbol{\lambda}_f)$  where  $\mathbf{\Lambda}_f = \text{diag}(\boldsymbol{\lambda}_f)$ . The gradient of  $L_1(\boldsymbol{\lambda}_f)$  with respect to  $\boldsymbol{\lambda}_f$ , denoted by  $\nabla L_1(\boldsymbol{\lambda}_f)$ , is easy to derive, which is omitted. Following the Armijo's rule [16], the search algorithm for  $\boldsymbol{\lambda}_f$  is as follows:

$$\boldsymbol{\lambda}_f^{(k+1)} = \boldsymbol{\lambda}_f^{(k)} + \beta^m \nabla L_1(\boldsymbol{\lambda}_f^{(k)}) \quad (15)$$

where  $m$  is the smallest integer satisfying

$$L_1(\boldsymbol{\lambda}_f^{(k+1)}) - L_1(\boldsymbol{\lambda}_f^{(k)}) > \sigma \beta^m \left\| \nabla L_1(\boldsymbol{\lambda}_f^{(k)}) \right\|^2 \quad (16)$$

$$P_f - \sum_i \lambda_{f,i}^{k+1} (\lambda_{h,i} \lambda_{x,i} + 1) > 0 \quad (17)$$

and  $0 < \sigma < 1$  and  $0 < \beta < 1$ . After convergence of the above search for a fixed  $t$ , a new search is started with an increased  $t$ . When  $1/t$  becomes small enough, the search for  $\mathbf{\Lambda}_f$  is considered completed for the given  $\mathbf{\Lambda}_x$ .

### B. Maximization of Sum Rate under Power Constraint and DPC (Algorithm 3)

ZFDPC is a scalar DPC, which is suboptimal compared to the vector DPC [12], [14], [17]. From now on, the vector DPC will be referred to as DPC. Given that the  $K$  users receive independent messages from the source, we can write the transmitted vector from the source as  $\mathbf{x} = \mathbf{x}_1 + \cdots + \mathbf{x}_K$  and its (source) covariance matrix as  $\mathbf{\Pi}_x = \mathbf{\Pi}_1 + \cdots + \mathbf{\Pi}_K$  where  $\mathbf{\Pi}_i$  is the covariance matrix of the signal

$\mathbf{x}_i$  meant for user  $i$ . Assuming the use of DPC in the descending order starting from user  $K$ , i.e., the interference from user  $j$  to user  $i$  for  $j > i$  is virtually absent, the achievable data rate for user  $i$  in bits/s/Hz is given by

$$I_{d,i} = \log_2 \frac{\left| \mathbf{H}_i \mathbf{F} \mathbf{H} \left( \sum_{j=1}^i \mathbf{\Pi}_j \right) \mathbf{H}^H \mathbf{F}^H \mathbf{H}_i^H + \mathbf{H}_i \mathbf{F} \mathbf{F}^H \mathbf{H}_i^H + \mathbf{I} \right|}{\left| \mathbf{H}_i \mathbf{F} \mathbf{H} \left( \sum_{j=1}^{i-1} \mathbf{\Pi}_j \right) \mathbf{H}^H \mathbf{F}^H \mathbf{H}_i^H + \mathbf{H}_i \mathbf{F} \mathbf{F}^H \mathbf{H}_i^H + \mathbf{I} \right|} \quad (18)$$

With any given set of the source covariance matrices  $\mathbf{\Pi}_i$  for  $i = 1, 2, \dots, K$ , a complete design of the vector DPC to achieve the rates in (18) can be made by following [17].

In the absence of total power constraint, the maximum possible data rate for user  $i$  is independent of  $\mathbf{\Pi}_j$  for  $j > i$  because of DPC. We can formulate the following problem:

$$\max_{\mathbf{A}_f, \mathbf{A}_x} \quad R_{sum,d} \doteq \sum_i^{KN} I_{d,i} \quad (19)$$

$$s.t. \quad tr\{\mathbf{\Pi}_x\} \leq P_x \quad (20)$$

$$tr\{\mathbf{F}(\mathbf{H}\mathbf{\Pi}_x\mathbf{H}^H + \mathbf{I})\mathbf{F}^H\} \leq P_f \quad (21)$$

A joint gradient search of  $\mathbf{F}, \mathbf{\Pi}_1, \dots, \mathbf{\Pi}_K$  can be performed directly to maximize the following penalized function of (19):

$$L_2(\mathbf{F}, \mathbf{A}_1, \dots, \mathbf{A}_K) \doteq \sum_i^{KN} I_{d,i} + \frac{1}{t_1} \log(P_x - tr\{\mathbf{\Pi}_x\}) + \frac{1}{t_2} \log(P_f - tr\{\mathbf{F}(\mathbf{H}\mathbf{\Pi}_x\mathbf{H}^H + \mathbf{I})\mathbf{F}^H\}) \quad (22)$$

where  $\mathbf{A}_i$  is such that  $\mathbf{\Pi}_i = \mathbf{A}_i \mathbf{A}_i^H$ . We can denote all parameters in  $\mathbf{F}, \mathbf{A}_1, \dots, \mathbf{A}_K$  by a single vector  $\mathbf{p}$ , and the gradient of  $L_2$  with respect to  $\mathbf{p}$  by  $\nabla L_2(\mathbf{p})$ . Similar to the case of (14), there are two loops in the search. The inner loop is for a fixed pair of  $(t_1, t_2)$  where the Armijo gradient search is conducted until the norm of  $\nabla L_2(\mathbf{p})$  is small enough. The outer loop corresponds to the increase of  $(t_1, t_2)$  until they are large enough.

To show an explicit expression of  $\nabla L_2(\mathbf{p})$ , it suffices to derive explicit expressions of  $\frac{\partial L_2}{\partial \mathbf{F}}$  and  $\frac{\partial L_2}{\partial \mathbf{A}_i}$  as follows. Following the rules of matrix differentials [18], we can show

$$\frac{\partial L_2}{\partial \mathbf{F}} = \sum_i^{KN} \frac{\partial I_{d,i}}{\partial \mathbf{F}} - \frac{2}{t_2} \frac{\mathbf{F}(\mathbf{H}\mathbf{\Pi}_x\mathbf{H}^H + \mathbf{I})}{P_f - tr\{\mathbf{F}(\mathbf{H}\mathbf{\Pi}_x\mathbf{H}^H + \mathbf{I})\mathbf{F}^H\}} \quad (23)$$

$$\frac{\partial L_2}{\partial \mathbf{A}_i} = \sum_i^{KN} \frac{\partial I_{d,i}}{\partial \mathbf{A}_j} - \frac{2}{t_1} \frac{\mathbf{A}_j}{P_x - tr\{\mathbf{\Pi}_x\}} - \frac{2}{t_2} \frac{\mathbf{H}^H \mathbf{F}^H \mathbf{F} \mathbf{H} \mathbf{A}_j}{P_f - tr\{\mathbf{F}(\mathbf{H}\mathbf{\Pi}_x\mathbf{H}^H + \mathbf{I})\mathbf{F}^H\}} \quad (24)$$

where the derivative of  $L_2$  with respect to the complex matrix  $\mathbf{F}$  is defined as  $\frac{\partial L_2}{\partial \mathbf{F}} = \frac{\partial L_2}{\partial Re\{\mathbf{F}\}} + j \frac{\partial L_2}{\partial Im\{\mathbf{F}\}}$ , and the same applies to  $\frac{\partial L_2}{\partial \mathbf{A}_j}$ . To derive  $\frac{\partial I_{d,i}}{\partial \mathbf{F}}$  and  $\frac{\partial I_{d,i}}{\partial \mathbf{A}_j}$ , we first define  $\mathbf{X}_i$  and  $\mathbf{Y}_i$  according to (18) such that  $I_{d,i} = \log_2 \frac{|\mathbf{X}_i|}{|\mathbf{Y}_i|}$ . Then, using  $\partial \log |\mathbf{X}| = tr\{\mathbf{X}^{-1} \partial \mathbf{X}\}$  [18], we have  $\partial I_{d,i} =$

$(\log_2 e) \text{tr} \{ \mathbf{X}_i^{-1} \partial \mathbf{X}_i - \mathbf{Y}_i^{-1} \partial \mathbf{Y}_i \}$ . It is easy to derive the differentials of  $\mathbf{X}_i$  and  $\mathbf{Y}_i$  with respect to the matrix  $\mathbf{F}$ . Applying the resulting expressions into the above expression, it follows that

$$\partial I_{d,i} = 2(\log_2 e) \text{Re} \left( \text{tr} \{ \mathbf{H}_i^H \mathbf{X}_i^{-1} \mathbf{M}_i \partial \mathbf{F}^H - \mathbf{H}_i^H \mathbf{Y}_i^{-1} \mathbf{N}_i \partial \mathbf{F}^H \} \right) \quad (25)$$

where  $\mathbf{M}_i = \mathbf{H}_i \mathbf{F} \mathbf{H} \left( \sum_{j=1}^i \mathbf{\Pi}_j \right) \mathbf{H}^H + \mathbf{H}_i \mathbf{F}$  and  $\mathbf{N}_i = \mathbf{H}_i \mathbf{F} \mathbf{H} \left( \sum_{j=1}^{i-1} \mathbf{\Pi}_j \right) \mathbf{H}^H + \mathbf{H}_i \mathbf{F}$ , and therefore

$$\frac{\partial I_{d,i}}{\partial \mathbf{F}} = 2(\log_2 e) \left( \mathbf{H}_i^H \mathbf{X}_i^{-1} \mathbf{M}_i - \mathbf{H}_i^H \mathbf{Y}_i^{-1} \mathbf{N}_i \right) \quad (26)$$

Similarly, one can verify that for  $j \leq i - 1$ ,

$$\frac{\partial I_{d,i}}{\partial \mathbf{A}_j} = 2(\log_2 e) \left( \mathbf{H}^H \mathbf{F}^H \mathbf{H}_i^H (\mathbf{X}_i^{-1} - \mathbf{Y}_i^{-1}) \mathbf{H}_i \mathbf{F} \mathbf{H} \mathbf{A}_j \right) \quad (27)$$

and for  $j = i$ ,  $\frac{\partial I_{d,i}}{\partial \mathbf{A}_j} = 2(\log_2 e) \left( \mathbf{H}^H \mathbf{F}^H \mathbf{H}_i^H \mathbf{X}_i^{-1} \mathbf{H}_i \mathbf{F} \mathbf{H} \mathbf{A}_j \right)$ , and for  $j > i$ ,  $\frac{\partial I_{d,i}}{\partial \mathbf{A}_j} = 0$ . The above algorithm is referred to as Algorithm 3.

### C. Minimization of Power under Rate Constraint (Algorithms 4-5)

We now address minimization of power under rate constraint. The total power consumed by the source and the relay is

$$P_d \doteq \text{tr} \{ \mathbf{\Pi}_x \} + \text{tr} \{ \mathbf{F} (\mathbf{H} \mathbf{\Pi}_x \mathbf{H}^H + \mathbf{I}) \mathbf{F}^H \} \quad (28)$$

Our problem now is to minimize the total power consumption subject to rate constraints:

$$\min_{\mathbf{F}, \mathbf{\Pi}_1, \dots, \mathbf{\Pi}_K} P_d \quad (29)$$

$$\text{subject to} \quad I_{d,i} \geq R_i \quad \forall i \in \{1, 2, \dots, K\} \quad (30)$$

where  $R_i$  is a desired data rate for user  $i$  in bits/s/Hz. To solve this problem, we can search for the optimal relay matrix  $\mathbf{F}$  and the optimal source covariance matrices  $\mathbf{\Pi}_1, \dots, \mathbf{\Pi}_K$  in an alternate fashion, where each cycle of the alternation is shown below.

1) *Source optimization with fixed  $\mathbf{F}$* : We now assume a fixed  $\mathbf{F}$  and present an algorithm for computing the optimal  $\mathbf{\Pi}_1, \dots, \mathbf{\Pi}_K$ . We will use the property that  $I_{d,i}$  is independent of  $\mathbf{\Pi}_{i+1}, \dots, \mathbf{\Pi}_K$  and is a concave function of  $\mathbf{\Pi}_i$ , and  $P$  is a linear function of  $\mathbf{\Pi}_1, \dots, \mathbf{\Pi}_K$ . It follows from (28) that

$$P_d = \sum_{i=1}^K \text{tr} \{ \mathbf{Q}_i \} + \text{tr} \{ \mathbf{F} \mathbf{F}^H \} \quad (31)$$

where  $\mathbf{Q}_i = (\mathbf{I} + \mathbf{H}^H \mathbf{F}^H \mathbf{F} \mathbf{H})^{H/2} \mathbf{\Pi}_i (\mathbf{I} + \mathbf{H}^H \mathbf{F}^H \mathbf{F} \mathbf{H})^{1/2}$ , and we have applied  $\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{BA})$ . Clearly,  $\mathbf{Q}_i$  and  $\mathbf{\Pi}_i$  are one-to-one mappings of each other. We now define

$$\mathbf{G}_i = \mathbf{H}_i \mathbf{F} \mathbf{H} (\mathbf{I} + \mathbf{H}^H \mathbf{F}^H \mathbf{F} \mathbf{H})^{-H/2} \quad (32)$$

$$\mathbf{S}_i = \mathbf{G}_i^H \left( \mathbf{G}_i \sum_{j=1}^{i-1} \mathbf{Q}_j \mathbf{G}_i^H + \mathbf{H}_i \mathbf{F} \mathbf{F}^H \mathbf{H}_i^H + \mathbf{I} \right)^{-1} \mathbf{G}_i \quad (33)$$

where  $\mathbf{S}_i$  depends on  $\mathbf{Q}_1, \dots, \mathbf{Q}_{i-1}$  but not any of  $\mathbf{Q}_i, \dots, \mathbf{Q}_K$ . Then, it follows from (18) that

$$I_{d,i} = \log_2 |\mathbf{S}_i \mathbf{Q}_i + \mathbf{I}| \quad (34)$$

where we have applied  $\log |\mathbf{AB} + \mathbf{I}| = \log |\mathbf{BA} + \mathbf{I}|$  with  $\mathbf{AB}$  being conjugate symmetric.

Based on (31) and (34), the optimal solution to the problem (29) for  $\mathbf{Q}_i$ , conditional upon  $\mathbf{F}, \mathbf{Q}_1, \dots, \mathbf{Q}_{i-1}$ , is given by the standard water filling solution. Namely, if the eigenvalue decomposition of  $\mathbf{S}_i$  is denoted by  $\mathbf{S}_i = \sum_{l=1}^r \lambda_{i,l} \mathbf{u}_{i,l} \mathbf{u}_{i,l}^H$  where  $\lambda_{i,l} > 0$ , then the optimal choice of  $\mathbf{Q}_i$  is  $\mathbf{Q}_i = \sum_{l=1}^r (v_i - \frac{1}{\lambda_{i,l}})^+ \mathbf{u}_{i,l} \mathbf{u}_{i,l}^H$  where  $(x)^+ = \max(x, 0)$  and  $v_i$  is such that  $I_{d,i} = R_i$ . (Note: In order to keep the solution inside the interior feasible region to ensure a good convergence behavior, we should choose  $I_{d,i}$  slightly larger than  $R_i$ .) Furthermore, with a fixed  $\mathbf{F}$ , the optimal solution for  $\mathbf{Q}_1, \dots, \mathbf{Q}_K$  (and hence  $\mathbf{\Pi}_1, \dots, \mathbf{\Pi}_K$ ) can be obtained one at a time sequentially by starting with  $\mathbf{\Pi}_1$ .

2) *Relay optimization with fixed  $\mathbf{\Pi}_1, \dots, \mathbf{\Pi}_K$* : We now assume that  $\mathbf{\Pi}_1, \dots, \mathbf{\Pi}_K$  are fixed. To find the optimal  $\mathbf{F}$ , we can use the gradient method to minimize the following penalized cost of (29):

$$L_3 = P_d - \frac{1}{t} \sum_i \log(I_{d,i} - R_i) \quad (35)$$

where the second term is the barrier, and both  $P_d$  and  $I_{d,i}$  are functions of  $\mathbf{F}$ . With the gradient  $\frac{\partial L_3}{\partial \mathbf{F}}$ , also denoted by  $\nabla L_3(\mathbf{F})$ , the Armijo search algorithm for the optimal  $\mathbf{F}$  is  $\mathbf{F}^{(k+1)} = \mathbf{F}^{(k)} - \beta^m \nabla L_3(\mathbf{F}^{(k)})$  where  $m$  is the smallest integer such that  $L_3(\mathbf{F}^{(k)}) - L_3(\mathbf{F}^{(k+1)}) > \sigma \beta^m \|\nabla L_3(\mathbf{F}^{(k)})\|^2$  and  $I_{d,i}(\mathbf{F}^{(k+1)}) - R_i > 0, \forall i \in \{1, \dots, K\}$  where  $0 < \beta < 1$  and  $0 < \sigma < 1$ . Note that the second condition  $I_{d,i}(\mathbf{F}^{(k+1)}) - R_i > 0$  is important to ensure that none of the rate constraints is violated. In fact, for good convergence behavior, for both the source optimization and the relay optimization, we need to keep  $\mathbf{F}, \mathbf{\Pi}_1, \dots, \mathbf{\Pi}_K$  strictly inside the interior feasible region of the problem.

The above algorithm for power minimization is referred to as Algorithm 4. Alternatively, we can solve the problem (29) by a joint gradient search similar to Algorithm 3, which will be referred to as Algorithm 5.

#### IV. MULTIUSER MIMO UPLINK RELAY

A multiuser MIMO uplink relay system is illustrated in Fig. 2, where we denote by  $\mathbf{H}_i^H \in \mathcal{C}^{M \times N}$  the channel matrix from user  $i$  to the relay, and by  $\mathbf{H}^H \in \mathcal{C}^{M \times M}$  the channel matrix from the relay to the access point. Then, we write the received signal at relay as

$$\mathbf{y}_r = \sum_{i=1}^K \mathbf{H}_i^H \mathbf{x}_i + \mathbf{n}_r \quad (36)$$

where  $\mathbf{x}_i$  is the signal transmitted from user  $i$ , and  $\mathbf{n}_r$  is the white Gaussian noise at the relay. The signal transmitted from the relay is

$$\mathbf{r} = \mathbf{F}^H \mathbf{y}_r \quad (37)$$

where  $\mathbf{F}^H$  is the relay matrix. The signal received at the access point is

$$\begin{aligned} \mathbf{y}_u &= \mathbf{H}^H \mathbf{r} + \mathbf{n}_u \\ &= \mathbf{H}^H \mathbf{F}^H \sum_{i=1}^K \mathbf{H}_i^H \mathbf{x}_i + \mathbf{H}^H \mathbf{F}^H \mathbf{n}_r + \mathbf{n}_u \end{aligned} \quad (38)$$

where  $\mathbf{n}_u$  is the white Gaussian noise at the access point. We assume the use of successive interference cancellation (SIC) at the access point, starting from user  $K$ . This means that the interference from user  $i$  to user  $k$  for  $i > k$  is virtually absent, and hence the achievable data rate for user  $k$  is

$$I_{u,k} = \log_2 \frac{\left| \mathbf{H}^H \mathbf{F}^H \left( \sum_{i=1}^k \mathbf{H}_i^H \mathbf{\Pi}_i \mathbf{H}_i \right) \mathbf{F} \mathbf{H} + \mathbf{H}^H \mathbf{F}^H \mathbf{F} \mathbf{H} + \mathbf{I} \right|}{\left| \mathbf{H}^H \mathbf{F}^H \left( \sum_{i=1}^{k-1} \mathbf{H}_i^H \mathbf{\Pi}_i \mathbf{H}_i \right) \mathbf{F} \mathbf{H} + \mathbf{H}^H \mathbf{F}^H \mathbf{F} \mathbf{H} + \mathbf{I} \right|} \quad (39)$$

where  $\mathbf{\Pi}_i = E\{\mathbf{x}_i \mathbf{x}_i^H\}$ .

##### A. Maximization of Sum Rate under Power Constraint (Algorithms 6-7)

The problem of maximizing the sum rate from all users under power constraints is formulated as follows:

$$\max_{\mathbf{F}, \mathbf{\Pi}_1, \dots, \mathbf{\Pi}_K} R_{sum,u} = \sum_{i=1}^K I_{u,i} = \log_2 \frac{\left| \mathbf{H}^H \mathbf{F}^H \left( \sum_{i=1}^K \mathbf{H}_i^H \mathbf{\Pi}_i \mathbf{H}_i \right) \mathbf{F} \mathbf{H} + \mathbf{H}^H \mathbf{F}^H \mathbf{F} \mathbf{H} + \mathbf{I} \right|}{\left| \mathbf{H}^H \mathbf{F}^H \mathbf{F} \mathbf{H} + \mathbf{I} \right|} \quad (40)$$

$$s.t. \quad tr\{\mathbf{\Pi}_i\} \leq P_i, \forall i \in \{1, 2, \dots, K\} \quad (41)$$

$$tr\left\{ \mathbf{F}^H \left( \sum_{i=1}^K \mathbf{H}_i^H \mathbf{\Pi}_i \mathbf{H}_i + \mathbf{I} \right) \mathbf{F} \right\} \leq P_f \quad (42)$$

Note that the sum rate of the uplink case is independent of the order of SIC, which is unlike the sum rate of the downlink case with DPC. To solve this problem, we can optimize each of  $\mathbf{F}, \mathbf{\Pi}_1, \dots, \mathbf{\Pi}_K$  in a cyclic fashion. The basic components in each cycle are shown below.

1) *Source optimization with fixed relay and other sources:* If all  $\mathbf{F}, \mathbf{\Pi}_1, \dots, \mathbf{\Pi}_K$ , but  $\mathbf{\Pi}_i$ , are fixed, we can define  $c = \log_2 |\mathbf{H}^H \mathbf{F}^H \mathbf{F} \mathbf{H} + \mathbf{I}|$  and

$$\mathbf{G}_i = \mathbf{H}^H \mathbf{F}^H \left( \sum_{j=1, j \neq i}^K \mathbf{H}_j^H \mathbf{\Pi}_j \mathbf{H}_j \right) \mathbf{F} \mathbf{H} + \mathbf{H}^H \mathbf{F}^H \mathbf{F} \mathbf{H} + \mathbf{I}$$

which are independent of  $\mathbf{\Pi}_i$ . Then, we can write

$$\begin{aligned} R_{sum,u} &= \log_2 |\mathbf{G}_i + \mathbf{H}^H \mathbf{F}^H \mathbf{H}_i^H \mathbf{\Pi}_i \mathbf{H}_i \mathbf{F} \mathbf{H}| - c \\ &= \log_2 \left| \mathbf{I} + \mathbf{G}_i^{-1/2} \mathbf{H}^H \mathbf{F}^H \mathbf{H}_i^H \mathbf{\Pi}_i \mathbf{H}_i \mathbf{F} \mathbf{H} \mathbf{G}_i^{-H/2} \right| + \log_2 |\mathbf{G}_i| - c \end{aligned} \quad (43)$$

The power constraint (42) is equivalent to

$$\text{tr} \{ \mathbf{F}^H \mathbf{H}_i^H \mathbf{\Pi}_i \mathbf{H}_i \mathbf{F} \} \leq P_f - \text{tr} \left\{ \mathbf{F}^H \left( \sum_{j=1, j \neq i}^K \mathbf{H}_j^H \mathbf{\Pi}_j \mathbf{H}_j + \mathbf{I} \right) \mathbf{F} \right\}$$

It should be clear now that with respect to  $\mathbf{\Pi}_i$  alone, the problem (40) is equivalent to the convex problem (1) which is solvable by the GWF algorithm.

2) *Relay optimization with fixed sources:* If  $\mathbf{\Pi}_1, \dots, \mathbf{\Pi}_K$  are fixed, then the problem (40) with respect to  $\mathbf{F}$  alone is similar to a problem solved in [3], the solution of which is stated below. Define the SVD of  $\mathbf{H}$  as  $\mathbf{H} = \mathbf{U}_h \mathbf{\Sigma}_h \mathbf{V}_h^H$  where  $\mathbf{\Sigma}_h = \text{diag}(\sigma_1, \dots, \sigma_M)$  with descending diagonal order, and the EVD of  $\mathbf{R} = \sum_{i=1}^K \mathbf{H}_i^H \mathbf{\Pi}_i \mathbf{H}_i$  as  $\mathbf{R} = \mathbf{E}_r \mathbf{\Lambda}_r \mathbf{E}_r^H$  where  $\mathbf{\Lambda}_r = \text{diag}(\lambda_1, \dots, \lambda_M)$  with descending diagonal order. Then, the optimal structure of  $\mathbf{F}$  is given by

$$\mathbf{F} = \mathbf{E}_r \mathbf{\Sigma}_f \mathbf{U}_h^H \quad (44)$$

where  $\mathbf{\Sigma}_f = \text{diag}(f_1, \dots, f_M)^{1/2} \geq 0$  which are to be determined. With (44), the problem (40) becomes

$$\begin{aligned} \max_{f_1, \dots, f_M} \quad & R_{sum,u} = \sum_{i=1}^M \log_2 \frac{\sigma_i^2 \lambda_i f_i + \sigma_i^2 f_i + 1}{\sigma_i^2 f_i + 1} \\ \text{s.t.} \quad & \sum_{i=1}^M (\lambda_i + 1) f_i \leq P_f \text{ and } f_i \geq 0 \quad \forall i \end{aligned} \quad (45)$$

Then, by the KKT method [15], we have

$$f_i = \frac{1}{2\sigma_i^2(1 + \lambda_i)} \left[ \sqrt{\lambda_i^2 + 4\lambda_i \sigma_i^2 \mu} - \lambda_i - 2 \right]^+ \quad (46)$$

where  $\mu$  is such that

$$\sum_{i=1}^M \frac{1}{2\sigma_i^2} \left[ \sqrt{\lambda_i^2 + 4\lambda_i \sigma_i^2 \mu} - \lambda_i - 2 \right]^+ = P_f.$$

The above algorithm that searches for  $\mathbf{F}, \mathbf{\Pi}_1, \dots, \mathbf{\Pi}_K$  in a cyclic fashion is referred to as Algorithm 6. Note that each component in Algorithm 6 is a convex optimization. Alternatively, we can solve the

problem (40) by a joint gradient search over  $\mathbf{F}, \mathbf{\Pi}_1, \dots, \mathbf{\Pi}_K$  simultaneously, which will be referred to as Algorithm 7. The details of Algorithm 7 are omitted because of its similarity to other joint gradient search algorithms.

### B. Minimization of Power under Rate Constraint (Algorithms 8-9)

The total power consumption for the uplink case is:

$$P_u = \sum_{i=1}^K \text{tr}\{\mathbf{\Pi}_i\} + \text{tr}\left\{\mathbf{F}^H \left(\sum_{i=1}^K \mathbf{H}_i^H \mathbf{\Pi}_i \mathbf{H}_i + \mathbf{I}\right) \mathbf{F}\right\} \quad (47)$$

With the assumption of SIC, the individual rate  $I_{u,i}$  for user  $i$  is given by (39). Hence, the problem is formulated as:

$$\min_{\mathbf{F}, \mathbf{\Pi}_1, \dots, \mathbf{\Pi}_K} P_u \quad (48)$$

$$s.t. \quad I_{u,i} \geq R_i, \quad \forall i \in \{1, 2, \dots, K\} \quad (49)$$

The problem (48) can be solved by a joint gradient search algorithm (Algorithm 9) which is omitted, or an alternate optimization algorithm (Algorithm 8) as shown below.

1) *Source optimization with fixed relay:* Since the order of the SIC is from  $K$  to 1,  $I_{u,i}$  is independent of  $\mathbf{\Pi}_{i+1}, \dots, \mathbf{\Pi}_K$ , which is a property also shared in the downlink case. With fixed  $\mathbf{F}, \mathbf{\Pi}_1, \dots, \mathbf{\Pi}_{i-1}$ , the optimal  $\mathbf{\Pi}_i$  can be found by a convex optimization same as in section III-C1.

2) *Relay optimization with fixed sources:* Given  $\mathbf{\Pi}_1, \dots, \mathbf{\Pi}_K$ , the optimal  $\mathbf{F}$  can be found by the following gradient method. Define the following cost with a barrier:

$$L_4 = \text{tr}\left\{\mathbf{F}^H \left(\sum_{i=1}^K \mathbf{H}_i^H \mathbf{\Pi}_i \mathbf{H}_i + \mathbf{I}\right) \mathbf{F}\right\} - \frac{1}{t} \sum_i \log(I_{u,i} - R_i) \quad (50)$$

It follows that

$$\frac{\partial L_4}{\partial \mathbf{F}} = 2 \left(\sum_{i=1}^K \mathbf{H}_i^H \mathbf{\Pi}_i \mathbf{H}_i + \mathbf{I}\right) \mathbf{F} - \frac{1}{t} \sum_i \frac{1}{I_{u,i} - R_i} \frac{\partial I_{u,i}}{\partial \mathbf{F}} \quad (51)$$

To derive  $\frac{\partial I_{u,i}}{\partial \mathbf{F}}$ , we first rewrite (39) as  $I_{u,i} = \log_2 \frac{|\mathbf{W}_i|}{|\mathbf{W}_{i-1}|}$ . Similar to the derivation of (26), it can be shown that

$$\frac{\partial I_{u,i}}{\partial \mathbf{F}} = 2(\log_2 e) (\mathbf{C}_i \mathbf{W}_i^{-1} \mathbf{H}^H - \mathbf{C}_{i-1} \mathbf{W}_{i-1}^{-1} \mathbf{H}^H) \quad (52)$$

where  $\mathbf{C}_i = \left(\mathbf{I} + \sum_{j=1}^i \mathbf{H}_j^H \mathbf{\Pi}_j \mathbf{H}_j\right) \mathbf{F} \mathbf{H}$ . The rest of the algorithm is the same as in section III-C2.

## V. MULTI-CARRIER EXTENSIONS

In the previous sections, we have assumed that there is a single carrier for power allocation. If one wants to use  $M_c$  (orthogonal) carriers for joint power allocation, the previously shown algorithms are also applicable after the following changes of notations are adopted.

For the downlink case, the signal models shown in (4)-(6) hold except that

$$\mathbf{x} = [\mathbf{x}(1)^T, \dots, \mathbf{x}(M_c)^T]^T \in \mathcal{C}^{MM_c \times 1} \quad (53)$$

$$\mathbf{y} = [\mathbf{y}(1)^T, \dots, \mathbf{y}(M_c)^T]^T \in \mathcal{C}^{MM_c \times 1} \quad (54)$$

$$\mathbf{n} = [\mathbf{n}(1)^T, \dots, \mathbf{n}(M_c)^T]^T \in \mathcal{C}^{MM_c \times 1} \quad (55)$$

$$\mathbf{H} = \text{diag}[\mathbf{H}(1), \dots, \mathbf{H}(M_c)] \in \mathcal{C}^{MM_c \times MM_c} \quad (56)$$

$$\mathbf{r} = [\mathbf{r}(1)^T, \dots, \mathbf{r}(M_c)^T]^T \in \mathcal{C}^{MM_c \times 1} \quad (57)$$

$$\mathbf{y}_i = [\mathbf{y}_i(1)^T, \dots, \mathbf{y}_i(M_c)^T]^T \in \mathcal{C}^{NM_c \times 1} \quad (58)$$

$$\mathbf{n}_i = [\mathbf{n}_i(1)^T, \dots, \mathbf{n}_i(M_c)^T]^T \in \mathcal{C}^{NM_c \times 1} \quad (59)$$

$$\mathbf{H}_i = \text{diag}[\mathbf{H}_i(1), \dots, \mathbf{H}_i(M_c)] \in \mathcal{C}^{NM_c \times MM_c} \quad (60)$$

and  $\mathbf{F} \in \mathcal{C}^{MM_c \times MM_c}$ , where for example  $\mathbf{x}(m)$  denotes the signal transmitted from the access point on the  $m$ th carrier. Note that the optimal  $\mathbf{F}$  is not necessarily block diagonal. In other words, the relay may use a different carrier to forward a stream of data that was received by the relay on another carrier [10]. Good (if not globally optimal) choices of  $\mathbf{F}$  along with the source covariance matrices at all carriers can be determined by any of the power allocation algorithms. For the uplink case, the signal models shown in (36)-(38) also hold after a similar change of definitions of the notations.

These notational changes do not affect any of the algorithms shown in this paper as long as the power constraint is for the sum power over all carriers and the rate of interest is also the sum rate over all carriers. However, the complexity of these algorithms will increase because of the increased dimensions.

## VI. SIMULATION RESULTS

For convenience of reference, all algorithms presented in Sections III and IV are summarized in Table I. For the simulation examples shown below, a sample set of computational times of all algorithms for a random channel realization and a random initialization are listed in the last line in Table I. All algorithms have roughly the same speed except Algorithm 1 which uses CVX and is much slower than others for a single run. Algorithm 1 uses geometric programming as proposed in [7], for which the GWF is not

applicable. However, unlike other algorithms, Algorithm 1 is globally convergent and needs no multiple runs associated with multiple initializations. When multiple runs are considered for other algorithms, they may become effectively slower than Algorithm 1. However, one can use the result from Algorithm 1 (for down link only) as an initialization for Algorithm 2 for a new research, which will be further discussed later.

Next, we show simulation examples to compare these algorithms. We assume that there are two users  $K = 2$ , each user is equipped with two antennas  $N = 2$ , the relay and the access point are both equipped with four antennas  $M = 4$ . A single carrier is assumed. Each of the channel parameters is realized independently using a complex Gaussian distribution with zero mean and unit variance. As assumed throughout this paper, every entry of the noise vectors has zero mean and unit variance. The performance in terms of either the sum rate or the total power is based on an average over 50 channel realizations. Our experience with 100 or more channel realizations did not lead to any significant change of results. Unless mentioned otherwise, the search conducted by each algorithm (except Algorithm 1 which is globally convergent) was initialized randomly, 20 random initializations were chosen for each realization of channel matrices, and the best result from the 20 initializations were selected for computing the performance. We have found that the performance difference between the “best” and “worst” from 20 initializations can be up to 20%. In general, the more initializations are used, the better is the chance the optimal solution is found. But the computational cost increases as the number of initialization increases.

Figure 3 compares the averaged sum rates achieved by the downlink Algorithms 1-3 versus the relay power  $P_f$ . The power at the source is fixed at  $P_x = 1$ . Algorithm 1 is based on the geometric programming proposed in [7]. Both Algorithms 1-2 are based on ZFDPC while Algorithm 3 is based on DPC. For Algorithm 2, there are two curves in this figure. For the lower curve, we used the results from Algorithm 1 as initializations for Algorithm 2. For the upper curve, we used random initializations. We see that except for the region of small relay power, Algorithm 1 yielded the least sum rate among the three algorithms while Algorithm 3 yielded the largest sum rate. In theory, Algorithm 3 should yield the largest sum rate for the entire region of relay power if a global optimum (including the optimal ordering of the DPC) is achieved. This figure suggests that in the small relay power region, Algorithm 3 was trapped in unfavorable local minima. Since ZFDPC and DPC are different coding schemes, the results from Algorithms 1-2 cannot unfortunately be used as good initializations for Algorithm 3. The complexity of DPC is much more complex than that of ZFDPC.

Figure 4 compares the averaged total power consumption required by the downlink Algorithms 4-5 versus individual rate constraint. Also shown in this figure is the power consumption based on the identity

relay matrix, i.e.,  $\mathbf{F} = \mathbf{I}$ , while the source covariance matrix is optimized by the source optimization subroutine in Algorithm 4. Algorithm 4 uses cyclic search while Algorithm 5 uses joint gradient search. The search directions for cyclic search are more limited than the joint gradient search. We see that when the data rate is high, the difference of power consumption is very large. The power consumption from Algorithm 5 is the least, i.e., the best.

Figure 5 compares the averaged sum rates achieved by the uplink Algorithms 6-7 versus the power constraint at the relay. The source power is fixed at  $P_i = 1$  for all  $i$ . It turns out that the two algorithms yield the same results. The relay optimization and the source optimization in Algorithm 6 (which is cyclic) are both convex, and Algorithm 7 uses the joint gradient search. The lower curve in this figure is based on the identity relay matrix, i.e.,  $\mathbf{F} = \mathbf{I}$ , while the source covariance matrices of all users are optimized by the source optimization subroutine in Algorithm 6.

Figure 6 compares the averaged total power consumption required by the uplink Algorithms 8-9 versus a common data rate of all users. Also shown in this figure is a curve based on the identity relay matrix, i.e.,  $\mathbf{F} = \mathbf{I}$ , while the source covariance matrices of all users are optimized by the source optimization subroutine in Algorithm 8. In this case, the joint gradient search by Algorithm 9 yields better results than the cyclic search by Algorithm 8.

Finally, Figure 7 illustrates an effect of joint multi-carrier power allocation. Here, the relay system is for downlink, there are two users ( $K = 2$ ), each user has two antennas ( $N = 2$ ), there are four antennas at the relay node and four antennas at the access point ( $M = 4$ ), and there are two carriers ( $M_c = 2$ ). For each of the two carriers, an independent channel realization was made. The first top curve is the sum rate over two users and two carriers, which was obtained by the joint multi-carrier power allocation. The second top curve is the sum rate over two users and two carriers, which was obtained by two separate single-carrier power allocations. The bottom two curves are the sum rates each summed over the two users for carrier 1 and carrier 2, respectively. The total power for the two carriers used for the first curve is twice that for each carrier used for the other curves. The power per carrier is the same for all curves. We see that there is an improvement of the sum rate by using joint multi-carrier power allocation, which is expected. However, the improvement is not large. It is known that the distribution of the singular values of a matrix of i.i.d random variables hardens (becomes invariant) as the dimension of the matrix increases [19]. Hence, if the number of antennas at each node becomes large, the improvement from the joint multi-carrier power allocation is expected to disappear.

## VII. CONCLUSION

In this paper, we have developed several computational strategies for a multiuser MIMO relay system where each node may be equipped with multiple antennas. The complexities of these algorithms are about the same, but their performances can be very much different. Although the central problem is non-convex, the joint gradient search for the relay matrix and the source covariance matrices, with multiple random initializations, has consistently yielded the best result. The use of logarithmic barrier functions, which is a key approach of the interior-point optimization methods, has been very effective for constrained optimizations. But for one case, the cyclic (or alternating) search for the relay matrix and the source covariance matrices yielded results similar to those by the joint gradient search. The GWF algorithm shown in this paper is a faster alternative to the CVX algorithm (or package) to solve the convex problem (1). In applications with practical coding methods, the rate-versus-power model of each link may need to be revised with simple penalty factors while the power allocation algorithms shown in this paper are still applicable. This paper has shown that fast algorithms for power allocation are very important to achieve the full potentials of MIMO relay systems with multiple-antenna users.

## APPENDIX A

### PROOF OF THEOREM 1

For any  $\mathbf{Q} \geq 0$  (i.e., positive semi-definite), we can write  $\mathbf{Q} = \mathbf{A}\mathbf{A}^H$  where  $\mathbf{A}$  is a full column rank matrix. With respect to  $\mathbf{A}$ , we can write the following Lagrangian function of (1):

$$L = -\log |I + \mathbf{H}\mathbf{A}\mathbf{A}^H\mathbf{H}^H| + \sum_{i=1}^m \mu_i (tr \{ \mathbf{B}_i \mathbf{A} \mathbf{A}^H \mathbf{B}_i^H \} - P_i) \quad (61)$$

The gradient of  $L$  with respect to  $\mathbf{A}$  can be found by using  $\partial \log |\mathbf{X}| = tr(\mathbf{X}^{-1} \partial \mathbf{X})$ ,  $\partial(\mathbf{X}\mathbf{X}^H) = (\partial \mathbf{X})\mathbf{X}^H + \mathbf{X}\partial \mathbf{X}^H$  and other basic tools [18]. The result is

$$\frac{\partial L}{\partial \mathbf{A}^H} \doteq \frac{\partial L}{\partial Re(\mathbf{A})^T} - j \frac{\partial L}{\partial Im(\mathbf{A})^T} = -2\mathbf{A}^H \left( \mathbf{H}^H (\mathbf{I} + \mathbf{H}\mathbf{A}\mathbf{A}^H\mathbf{H}^H)^{-1} \mathbf{H} - \sum_{i=1}^m \mu_i \mathbf{B}_i^H \mathbf{B}_i \right) \quad (62)$$

Then, the complete K.K.T. conditions [15] of the problem (1) with respect to  $\mathbf{A}$  can be written as

$$-\mathbf{A}^H \left( \mathbf{H}^H (\mathbf{I} + \mathbf{H}\mathbf{A}\mathbf{A}^H\mathbf{H}^H)^{-1} \mathbf{H} - \sum_{i=1}^m \mu_i \mathbf{B}_i^H \mathbf{B}_i \right) = 0 \quad (63)$$

$$tr \{ \mathbf{B}_i \mathbf{A} \mathbf{A}^H \mathbf{B}_i^H \} - P_i \leq 0 \quad (64)$$

$$\mu_i \geq 0 \quad (65)$$

$$\mu_i (tr \{ \mathbf{B}_i \mathbf{A} \mathbf{A}^H \mathbf{B}_i^H \} - P_i) = 0 \quad (66)$$

where  $i = 1, \dots, m$ .

Although the problem (1) with respect to  $\mathbf{A}$  is not convex, we now show that the generalized KKT conditions [15] of the problem (1) with respect to  $\mathbf{Q} \geq 0$ , which is convex, are equivalent to (63)-(66). Consider  $L$  as in (61) with  $\mathbf{A}\mathbf{A}^H$  replaced by  $\mathbf{Q}$ . It follows that

$$\frac{\partial L}{\partial \mathbf{Q}} = -\mathbf{H}^H (\mathbf{I} + \mathbf{H}\mathbf{Q}\mathbf{H}^H)^{-1} \mathbf{H} + \sum_{i=1}^m \mu_i \mathbf{B}_i^H \mathbf{B}_i \quad (67)$$

We define a vector operator for a complex conjugate symmetric matrix as follows:

$$\text{vec}(\mathbf{Q}) \doteq \begin{bmatrix} \text{vec}(\text{Re}\{\mathbf{Q}\}) \\ \text{vec}(\text{Im}\{\mathbf{Q}\}) \end{bmatrix}$$

Here,  $\text{vec}(\text{Re}\{\mathbf{Q}\})$  stacks up all elements from  $\text{Re}\{\mathbf{Q}\}$ , and  $\text{vec}(\text{Im}\{\mathbf{Q}\})$  stacks up all elements from  $\text{Im}\{\mathbf{Q}\}$ . Assume  $\mathbf{Q} \in \mathcal{C}^{n \times n}$ . Then,  $\text{vec}(\mathbf{Q}) \in \mathcal{R}^{2n^2 \times 1}$ . Now, based on (5.95) in [15], we have the following sufficient generalized KKT conditions:

$$\text{vec} \left( -\mathbf{H}^H (\mathbf{I} + \mathbf{H}\mathbf{Q}\mathbf{H}^H)^{-1} \mathbf{H} + \sum_{i=1}^m \mu_i \mathbf{B}_i^H \mathbf{B}_i \right) - \boldsymbol{\omega} = 0 \quad (68)$$

$$\text{tr} \{ \mathbf{B}_i \mathbf{Q} \mathbf{B}_i^H \} - P_i \leq 0 \quad (69)$$

$$\mu_i \geq 0 \quad (70)$$

$$\mu_i (\text{tr} \{ \mathbf{B}_i \mathbf{Q} \mathbf{B}_i^H \} - P_i) = 0 \quad (71)$$

$$\boldsymbol{\omega}^T \text{vec}(\mathbf{Q}) = 0 \quad (72)$$

where  $i = 1, \dots, m$ ,  $\boldsymbol{\omega} \in \mathcal{R}^{2n^2 \times 1}$ . Also,  $\mathbf{Q} \in \mathcal{K} \doteq \{\mathbf{Q}' \mid \mathbf{Q}' \geq 0\}$ , and  $\boldsymbol{\omega}$  is in the dual cone of  $\mathcal{K}$ , i.e.,  $\boldsymbol{\omega} \in \mathcal{K}^D \doteq \{\boldsymbol{\omega} \mid \boldsymbol{\omega}^T \text{vec}(\mathbf{Q}') \geq 0 \ \forall \ \mathbf{Q}' \geq 0\}$ . The term  $-\boldsymbol{\omega}$  in (68) is due to the constraint  $-\mathbf{Q} \leq 0$ , for which we have used  $\frac{\partial \text{vec}^T(\mathbf{Q})}{\partial \text{vec}(\mathbf{Q})} = \mathbf{I}$ .

Note that for any two complex conjugate symmetric and positive semi-definite matrices  $\mathbf{A}'$  and  $\mathbf{B}'$ , the following equations are equivalent:  $\mathbf{A}'^H \mathbf{B}' = 0 \Leftrightarrow \text{tr}(\mathbf{A}'^H \mathbf{B}') = 0 \Leftrightarrow \text{Re}\{\mathbf{A}'\}^T \text{Re}\{\mathbf{B}'\} + \text{Im}\{\mathbf{A}'\}^T \text{Im}\{\mathbf{B}'\} = 0 \Leftrightarrow \text{vec}(\mathbf{A}')^T \text{vec}(\mathbf{B}') = 0$ . It is then easy to show, similar to Example 2.24 in [15], that  $\mathcal{K} = \mathcal{K}^D$ . Then, as long as  $\frac{\partial L}{\partial \mathbf{Q}} \geq 0$  and  $\mathbf{Q} = \mathbf{A}\mathbf{A}^H$ , we have that (68) implies  $\boldsymbol{\omega} \in \mathcal{K}^D$ , (63) implies (72) and vice versa. On the other hand, if  $\frac{\partial L}{\partial \mathbf{Q}} \geq 0$  does not hold, then  $\boldsymbol{\omega} \in \mathcal{K}^D$  does not hold because of (68). Therefore, if and only if  $\frac{\partial L}{\partial \mathbf{Q}} \geq 0$ , (63)-(66) are equivalent to (68)-(72).

Next, we construct an optimal structure of  $\mathbf{Q}$  based on (63). Since  $\mathbf{K}\mathbf{K}^H = \sum_{i=1}^m \mu_i \mathbf{B}_i^H \mathbf{B}_i$  and  $\mathbf{K}$  is non-singular, (63) is equivalent to

$$-\mathbf{A}^H \mathbf{K} \left( \mathbf{K}^{-1} \mathbf{H}^H (\mathbf{I} + \mathbf{H} \mathbf{K}^{-H} \mathbf{K}^H \mathbf{A} \mathbf{A}^H \mathbf{K} \mathbf{K}^{-1} \mathbf{H}^H)^{-1} \mathbf{H} \mathbf{K}^{-H} - \mathbf{I} \right) = 0 \quad (73)$$

Define the SVD of  $\mathbf{H}\mathbf{K}^{-H}$  as

$$\mathbf{H}\mathbf{K}^{-H} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H = \mathbf{U} \begin{pmatrix} \mathbf{\Sigma}_1 & \\ & \mathbf{\Sigma}_2 \end{pmatrix} \begin{pmatrix} \mathbf{V}_1 & \mathbf{V}_2 \end{pmatrix}^H \quad (74)$$

where  $\mathbf{U}$  and  $\mathbf{V}$  are square unitary matrices,  $\mathbf{\Sigma}_1$  (square) and  $\mathbf{\Sigma}_2$  (possibly non-square) are diagonal, all the diagonal elements of  $\mathbf{\Sigma}_1$  are larger than one, and all the diagonal elements of  $\mathbf{\Sigma}_2$  are less than or equal to one. We now assume that  $\mathbf{K}^H \mathbf{A} = \mathbf{V}_1 \mathbf{T}$  where  $\mathbf{T}$  is non-singular. Then, (73) is equivalent to the following:

$$\begin{aligned} & -\mathbf{A}^H \mathbf{K} \left( \mathbf{K}^{-1} \mathbf{H}^H (\mathbf{I} + \mathbf{H}\mathbf{K}^{-H} \mathbf{K}^H \mathbf{A} \mathbf{A}^H \mathbf{K} \mathbf{K}^{-1} \mathbf{H}^H)^{-1} \mathbf{H}\mathbf{K}^{-H} - \mathbf{I} \right) = 0 \\ \Leftrightarrow^{(a)} & -\mathbf{T}^H \mathbf{V}_1^H \left( \mathbf{V} \mathbf{\Sigma}^T \mathbf{U}^H (\mathbf{I} + \mathbf{U} \mathbf{\Sigma} \mathbf{V}^H \mathbf{V}_1 \mathbf{T} \mathbf{T}^H \mathbf{V}_1^H \mathbf{V} \mathbf{\Sigma}^T \mathbf{U}^H)^{-1} \mathbf{U} \mathbf{\Sigma} \mathbf{V}^H - \mathbf{I} \right) = 0 \\ \Leftrightarrow^{(b)} & -\mathbf{T}^H \left( \begin{pmatrix} \mathbf{\Sigma}_1 & \mathbf{0} \end{pmatrix} \left( \mathbf{I} + \begin{pmatrix} \mathbf{\Sigma}_1 \\ \mathbf{0} \end{pmatrix} \mathbf{T} \mathbf{T}^H \begin{pmatrix} \mathbf{\Sigma}_1 & \mathbf{0} \end{pmatrix} \right)^{-1} \mathbf{\Sigma} - \begin{pmatrix} \mathbf{I} & \mathbf{0} \end{pmatrix} \right) \mathbf{V}^H = 0 \\ \Leftrightarrow^{(c)} & -\mathbf{T}^H \left( \begin{pmatrix} \mathbf{\Sigma}_1 & \mathbf{0} \end{pmatrix} \left( \mathbf{I} - \begin{pmatrix} \mathbf{\Sigma}_1 \\ \mathbf{0} \end{pmatrix} ((\mathbf{T} \mathbf{T}^H)^{-1} + \mathbf{\Sigma}_1^2)^{-1} \begin{pmatrix} \mathbf{\Sigma}_1 & \mathbf{0} \end{pmatrix} \right) \mathbf{\Sigma} - \begin{pmatrix} \mathbf{I} & \mathbf{0} \end{pmatrix} \right) \mathbf{V}^H = 0 \\ \Leftrightarrow^{(d)} & -\mathbf{T}^H \left( \left( \begin{pmatrix} \mathbf{\Sigma}_1^2 & \mathbf{0} \end{pmatrix} - \mathbf{\Sigma}_1^2 ((\mathbf{T} \mathbf{T}^H)^{-1} + \mathbf{\Sigma}_1^2)^{-1} \begin{pmatrix} \mathbf{\Sigma}_1^2 & \mathbf{0} \end{pmatrix} \right) - \begin{pmatrix} \mathbf{I} & \mathbf{0} \end{pmatrix} \right) \mathbf{V}^H = 0 \\ \Leftrightarrow^{(e)} & \mathbf{\Sigma}_1^2 - \mathbf{\Sigma}_1^2 \left( (\mathbf{T} \mathbf{T}^H)^{-1} + \mathbf{\Sigma}_1^2 \right)^{-1} \mathbf{\Sigma}_1^2 - \mathbf{I} = 0 \\ \Leftrightarrow^{(f)} & \mathbf{\Sigma}_1^2 - \mathbf{\Sigma}_1^2 \left( \mathbf{\Sigma}_1^{-2} - \mathbf{\Sigma}_1^{-2} (\mathbf{T} \mathbf{T}^H + \mathbf{\Sigma}_1^{-2})^{-1} \mathbf{\Sigma}_1^{-2} \right) \mathbf{\Sigma}_1^2 - \mathbf{I} = 0 \\ \Leftrightarrow^{(g)} & \mathbf{T} \mathbf{T}^H = \mathbf{I} - \mathbf{\Sigma}_1^{-2} \end{aligned} \quad (75)$$

where for (c) and (f) we used the matrix inverse lemma. We see that since  $\mathbf{T} \mathbf{T}^H = \mathbf{I} - \mathbf{\Sigma}_1^{-2} > 0$ , the above solution for  $\mathbf{T}$ , and hence the corresponding  $\mathbf{A}$ , is a valid solution.

The above solution of  $\mathbf{K}^H \mathbf{A}$  has the same span as  $\mathbf{V}_1$ . A simple observation of the above analysis also suggests that as long as the span of  $\mathbf{K}^H \mathbf{A}$  belongs to that of  $\mathbf{V}_1$ , a matrix  $\mathbf{T}$  exists such that  $\mathbf{K}^H \mathbf{A} = \mathbf{V}_1' \mathbf{T}$  satisfies (73) where  $\mathbf{V}_1'$  is a sub-matrix (selected columns) of  $\mathbf{V}_1$ . On the other hand, if the span of  $\mathbf{K}^H \mathbf{A}$  contains a vector from  $\mathbf{V}_2$ , i.e.,  $\mathbf{K}^H \mathbf{A} = \mathbf{V}_2' \mathbf{T}$  where  $\mathbf{V}_2'$  has a column vector from  $\mathbf{V}_2$ , then there does not exist such a matrix  $\mathbf{T}$  for  $\mathbf{A}$  to satisfy (73), or equivalently the corresponding “solution”  $\mathbf{T} \mathbf{T}^H$  would be non-positive semi-definite which contradicts to the fundamental nature of  $\mathbf{T} \mathbf{T}^H$ . Therefore, the highest rank solution of  $\mathbf{A}$  to satisfy (73) is given by  $\mathbf{A} = \mathbf{K}^{-H} \mathbf{V}_1 \mathbf{T}$  where

$\mathbf{T} = (\mathbf{I} - \Sigma_1^{-2})^{1/2}$ . Equivalently, the highest rank solution of  $\mathbf{Q}$  to satisfy (73) is given by

$$\begin{aligned}\mathbf{Q} &= \mathbf{A}\mathbf{A}^H = \mathbf{K}^{-H}\mathbf{V}_1\mathbf{T}\mathbf{T}^H\mathbf{V}_1^H\mathbf{K}^{-1} \\ &= \mathbf{K}^{-H}\mathbf{V}_1(\mathbf{I} - \Sigma_1^{-2})\mathbf{V}_1^H\mathbf{K}^{-1} \\ &= \mathbf{K}^{-H}\mathbf{V}(\mathbf{I} - \Sigma^{-2})^+\mathbf{V}^H\mathbf{K}^{-1}\end{aligned}\tag{76}$$

where  $\Sigma^{-2} = (\Sigma^T \Sigma)^{-1}$ , the inverse of a zero (squared singular value) would be treated as positive infinity, and  $(\mathbf{I} - \Sigma^{-2})^+$  applies  $(x)^+ \doteq \max(x, 0)$  on each diagonal element of itself.

With (76) and (67), one can verify that

$$\frac{\partial L}{\partial \mathbf{Q}} = \mathbf{K}\mathbf{V} \left( \mathbf{I} - \Sigma^T \left( \mathbf{I} + \Sigma (\mathbf{I} - \Sigma^{-2})^+ \Sigma^T \right)^{-1} \Sigma \right) \mathbf{V}^H \mathbf{K}^H \geq 0 \tag{77}$$

Note that the  $i$ th diagonal element of the diagonal matrix between  $\mathbf{V}$  and  $\mathbf{V}^H$  in (77), denoted by  $d_i$ , is

$$\begin{aligned}d_i &= 1 - \sigma_i^2 (1 + \sigma_i^2 (1 - \sigma_i^{-2})^+)^{-1} \\ &= \begin{cases} 1 - \sigma_i^2 > 0 & \text{if } \sigma_i^2 < 1 \\ 0 & \text{if } \sigma_i^2 \geq 1 \end{cases}\end{aligned}\tag{78}$$

where  $\sigma_i$  is the  $i$ th diagonal element of  $\Sigma$ . If we did not use the highest rank solution for  $\mathbf{Q}$  as in (76), then there would be a  $d_i = 1 - \sigma_i^2 < 0$  associated with a  $\sigma_i^2 > 1$  and hence (77) would not hold and hence the corresponding  $\omega$  from (68) would not belong to  $\mathcal{K}^D$ .

With the optimal  $\mathbf{Q}$  given in (76), which is a function of  $\boldsymbol{\mu} = [\mu_1, \dots, \mu_m]$ , the remaining problem is to find the optimal  $\boldsymbol{\mu}$ . Since the effective KKT equations for  $\boldsymbol{\mu}$  are the same for both (63)-(66) and (68)-(72), the optimal  $\boldsymbol{\mu}$  can be found by using either the dual problem of (1) with respect to  $\mathbf{A}$  or the dual problem of (1) with respect to  $\mathbf{Q}$ . Choosing the former, we can find the optimal  $\boldsymbol{\mu}$  by solving (3). The dual problem of (1) with respect to  $\mathbf{Q}$  is the same as (3) except for the additional term  $-\text{vec}^T(\mathbf{Q})\omega$  which is however maximized to zero by  $\omega$  for any  $\boldsymbol{\mu}$ .

The proof of the theorem is completed. In the next section, we show how to find the optimal  $\boldsymbol{\mu}$  in more details. For the primal problem (1),  $\mathbf{Q}$  has  $2n^2$  real elements. (Even under the constraint  $\mathbf{Q} = \mathbf{Q}^H$ ,  $\mathbf{Q}$  has  $\frac{n(n+1)}{2}$  free real-part elements,  $\frac{n(n-1)}{2}$  free imaginary-part elements, and hence total  $n^2$  free real elements.) For the dual problem (3), there are  $m$  real variables in  $\boldsymbol{\mu}$ . If  $m < n^2$ , it is reasonable to expect the dual problem to be less costly to solve.

## APPENDIX B

### COMPUTATION OF THE DUAL PROBLEM IN THEOREM 1

Since the dual problem is convex, we can follow the interior-point method [15] and define the following dual function with logarithmic barrier terms:

$$D(\boldsymbol{\mu}) = -\log |\mathbf{I} + \mathbf{H}\mathbf{Q}(\boldsymbol{\mu})\mathbf{H}^H| + \sum_{i=1}^m \mu_i (\text{tr}(\mathbf{B}_i \mathbf{Q}(\boldsymbol{\mu}) \mathbf{B}_i^H) - P_i) + \frac{1}{t} \sum_i \log \mu_i \quad (79)$$

where we use  $\mathbf{Q}(\boldsymbol{\mu})$  to stress that  $\mathbf{Q}$  is a function of  $\boldsymbol{\mu}$ . Note that the first two terms in (79) equal to  $\min_{\mathbf{Q} \geq 0} L$ , which we want to maximize subject to  $\boldsymbol{\mu} \geq 0$ . For each choice of  $t$ , we can apply the Newton's method [15] to find the optimal  $\boldsymbol{\mu}$ , i.e.,

$$\boldsymbol{\mu}^{(k+1)} = \boldsymbol{\mu}^{(k)} + \gamma^{(k)} (\nabla^2 D(\boldsymbol{\mu}^{(k)}))^{-1} \nabla D(\boldsymbol{\mu}^{(k)}) \quad (80)$$

where  $k$  denotes the iteration index and the scalar  $\gamma^{(k)}$  is determined by the backtracking line search. Upon convergence for each  $t$ , we can increase  $t$  by a factor  $\delta > 1$  and continue a new cycle of the Newton's search. The above process continues until  $1/t$  is smaller than a pre-specified number  $\epsilon$ .

The computation of the gradient vector  $\nabla D(\boldsymbol{\mu}^{(k)})$  and the Hessian matrix  $\nabla^2 D(\boldsymbol{\mu}^{(k)})$  is straightforward although the detailed expressions are lengthy. Since  $\mathbf{Q}(\boldsymbol{\mu})$  depends on the eigenvalue decomposition of  $\mathbf{K}^{-1} \mathbf{H}^H \mathbf{H} \mathbf{K}^{-H}$  and the computation of  $\mathbf{K} = (\sum_{i=1}^m \mu_i \mathbf{B}_i^H \mathbf{B}_i)^{1/2}$  also needs the eigenvalue decomposition of  $\sum_{i=1}^m \mu_i \mathbf{B}_i^H \mathbf{B}_i$ , we need to use the first-order and second-order differentials of eigenvalues and eigenvectors. The basic formulas for these differentials can be found in [18]. The detailed expressions of the gradient and the Hessian are omitted to save space.

To avoid possible numerical problems in computing the differentials of eigenvectors when there are multiple identical eigenvalues, we added a small random perturbation matrix to  $\sum_{i=1}^m \mu_i \mathbf{B}_i^H \mathbf{B}_i$  in our program, which proved to be very effective. A complete Matlab script of the GWF algorithm is available at <http://www.ee.ucr.edu/~yhua/GWF.pdf>.

## APPENDIX C

### A COMPARISON OF GWF AND CVX

To show a comparison of our GWF algorithm with CVX in [13], we ran both algorithms on a desktop with 2.40GHz CPU. We chose  $P_1 = 1$ ,  $P_2 = 1.5$ ,  $\mathbf{B}_1 = \mathbf{I}$ , and used the complex Gaussian distribution

with zero mean and unit variance to randomly choose each element in the following matrices:

$$\mathbf{H} = \begin{pmatrix} -0.6705 + 0.3791i & 0.1469 + 0.4499i & -0.2913 - 0.3867i & 0.1568 - 0.0536i \\ 0.2398 - 0.3460i & -0.0702 - 1.0615i & -0.4482 + 0.0759i & -1.0125 + 0.5067i \\ -0.8170 + 0.3401i & -0.5652 + 0.1424i & 0.1243 - 0.1684i & 0.2645 - 0.2377i \\ -0.7213 - 0.5363i & -0.1463 - 0.3667i & -0.7448 + 0.4854i & 0.1717 + 0.0345i \end{pmatrix}$$

$$\mathbf{B}_2 = \begin{pmatrix} 0.1993 + 0.1027i & -0.6859 + 0.4280i & 0.1457 + 0.3800i & 0.2031 + 0.5548i \\ 0.5582 + 0.2944i & -0.3429 - 0.4255i & 0.5535 - 0.8565i & 0.6080 - 0.5549i \\ 0.3102 - 0.1320i & 0.1658 + 0.4059i & 0.1225 + 0.7685i & 0.7242 + 0.1927i \\ -0.1438 + 1.2477i & -0.4989 + 0.3501i & 0.0825 - 0.8049i & -0.5126 + 0.4826i \end{pmatrix}$$

For the GWF algorithm, the initial elements of  $\boldsymbol{\mu}^{(0)}$  were randomly chosen between zero and  $10^{-2}$ . We chose  $\nabla D(\boldsymbol{\mu})^T (\nabla^2 D(\boldsymbol{\mu}))^{-1} \nabla D(\boldsymbol{\mu}) < 10^{-2}$  as the stopping criterion for the inner loop (for fixed  $t$ ). We also chose  $t^{(1)} = 2$  and  $t^{(i+1)} = 2t^{(i)}$ , and finally  $2/t < 10^{-4}$  as the stopping criterion for the outer loop. We noticed that for each  $t$ , the inner loop converged after about 8 iterations.

At the convergence, the following results from the GWF algorithm and the CVX algorithm were obtained:

$$\mathbf{Q}_{GWF} = \begin{pmatrix} 0.3726 & 0.1804 - 0.0634i & 0.0470 - 0.0795i & -0.1740 - 0.0078i \\ 0.1804 + 0.0634i & 0.2722 & -0.0779 - 0.1381i & -0.1265 - 0.1644i \\ 0.0470 + 0.0795i & -0.0779 + 0.1381i & 0.1643 & 0.0893 + 0.0208i \\ -0.1740 + 0.0078i & -0.1265 + 0.1644i & 0.0893 - 0.0208i & 0.1909 \end{pmatrix}$$

$$\mathbf{Q}_{CVX} = \begin{pmatrix} 0.3726 & 0.1804 - 0.0634i & 0.0469 - 0.0796i & -0.1739 - 0.0078i \\ 0.1804 + 0.0634i & 0.2722 & -0.0779 - 0.1382i & -0.1265 - 0.1644i \\ 0.0469 + 0.0796i & -0.0779 + 0.1382i & 0.1643 & 0.0894 + 0.0208i \\ -0.1739 + 0.0078i & -0.1265 + 0.1644i & 0.0894 - 0.0208i & 0.1909 \end{pmatrix}$$

These two matrices agree with each other very well. Both GWF and CVX achieve the same value of capacity 2.6139 in bits/s/Hz (i.e.,  $-J$  in (1)). But GWF took 3.40 seconds while CVX took 14.94 seconds. GWF is about four times faster than CVX. Note that the dimension of  $\mathbf{Q}$  used here is larger than that used for Algorithms 2 and 6 shown in Table I

Figure 8 shows how  $\boldsymbol{\mu}$  of the GWF converged to the optimal as the outer iterations continued. We see that  $\mu_2$  approaches to zero, which means that the second power constraint is satisfied automatically while the first power constraint is active. Figure 9 illustrates the capacity ( $-J$ ) as function of the barrier constant  $t$ .

## REFERENCES

- [1] X. Tang and Y. Hua, "Optimal design of non-regenerative MIMO wireless relay," *IEEE Transactions on Wireless Communications*, vol. 6, pp. 1398–1407, April 2007.
- [2] O. Munoz-Medina, J. Vidal, and A. Agustin, "Linear transceiver design in nonregenerative relays with channel state information," *IEEE Transactions on Signal Processing*, vol. 55, no. 6, pp. 2593–2604, June 2007.
- [3] Z. Fang, Y. Hua, and J. Koshy, "Joint source and relay optimization for a non-regenerative MIMO relay," in *IEEE Workshop on Sensor Array and Multi-channel Processing*, Waltham, MA, July 2006.
- [4] Y. Fan and J. Thompson, "MIMO configurations for relay channels: Theory and practice," *IEEE Transactions on Wireless Communications*, vol. 6, pp. 1774–1786, May 2007.
- [5] Y. Rong, X. Tang, and Y. Hua, "A unified framework for optimizing linear non-regenerative multicarrier MIMO relay communication systems," *IEEE Transactions on Signal Processing*, vol. 57, no. 12, pp. 4837–4852, Dec 2009.
- [6] Y. Rong and Y. Hua, "Optimality of diagonalization of multi-hop MIMO relays," *IEEE Transactions on Wireless Communications*, to appear.
- [7] C. Chae, T. Tang, R. Heath, and S. Cho, "MIMO relaying with linear processing for multiuser transmission in fixed relay networks," *IEEE Transactions on Signal Processing*, vol. 56, pp. 727–738, Feb. 2008.
- [8] L. Weng and R. Murch, "Multi-user MIMO relay system with self-interference cancellation," in *IEEE WCNC*, 2007.
- [9] S. Jafar, K. Gomadam, and C. Huang, "Duality and rate optimization for multiple access and broadcast channels with amplify-and-forward relays," *IEEE Transactions on Information Theory*, vol. 53, pp. 3350–3370, Oct. 2007.
- [10] W. Zhang and U. Mitra, "Channel-adaptive frequency-domain relay processing in multicarrier multihop transmission," in *ICASSP 2008*, April 2008, Las Vegas, NV.
- [11] S. Berger, M. Kuhn, A. Wittneben, T. Unger, and A. Klein, "Recent advances in amplify-and-forward two-hop relaying," *IEEE Communications Magazine*, pp. 50–56, July 2009.
- [12] S. Vishwanath, M. Jindal, and A. Goldsmith, "Duality, achievable rates and sum-rate capacity of Gaussian MIMO broadcast channels," *IEEE Transactions on Information Theory*, vol. 49, pp. 2658–2668, Oct. 2003.
- [13] M. Grant and S. Boyd, *CVX: Matlab software for disciplined convex programming*, <http://stanford.edu/~boyd/cvx>, 2008.
- [14] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," *IEEE Transactions on Information Theory*, vol. 52, pp. 3936–3964, Sept. 2006.
- [15] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.
- [16] D. Bertsekas, *Nonlinear Programming*, Athena Scientific, second edition, 1995.
- [17] S.-C. Lin and H.-J. Su, "Practical vector dirty paper coding for MIMO gaussian broadcast channels," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 7, pp. 1345–1357, Sept. 2007.
- [18] J. Magnus and H. Neudecker, *Matrix Differential Calculus with Applications in Statistics and Econometrics*, John Wiley & Sons, 1999.
- [19] R. Wegmann, "The asymptotic eigenvalue-distribution for a certain class of random matrices," *Journal of Mathematical Analysis and Applications*, vol. 56, pp. 113–132, 1976.

TABLE I

SUMMARY OF POWER ALLOCATION ALGORITHMS FOR A MULTIUSER MIMO RELAY SYSTEM. THE SAMPLE RUN TIMES WERE BASED ON A DESKTOP WITH 2.40GHz CPU, TWO USERS EACH WITH TWO ANTENNAS AND A RELAY WITH FOUR ANTENNAS.

	Alg. 1	Alg. 2	Alg. 3	Alg. 4	Alg. 5	Alg. 6	Alg. 7	Alg. 8	Alg. 9
Section No.	III-A	III-A	III-B	III-C	III-C	IV-A	IV-A	IV-B	IV-B
Downlink	✓	✓	✓	✓	✓				
Uplink						✓	✓	✓	✓
Max Rate	✓	✓	✓			✓	✓		
Min Power				✓	✓			✓	✓
ZFDPC	✓	✓							
DPC			✓	✓	✓				
SIC						✓	✓	✓	✓
Cyclic Search		✓		✓		✓		✓	
Joint Search	✓		✓		✓		✓		✓
Use of GWF		✓				✓			
Sample Run Time in Sec	17.10	5.12	4.38	7.44	6.32	8.15	6.91	4.18	3.92

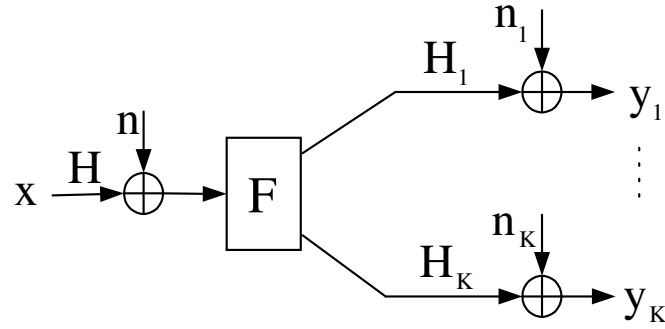


Fig. 1. Diagram of a multiuser MIMO relay downlink system.

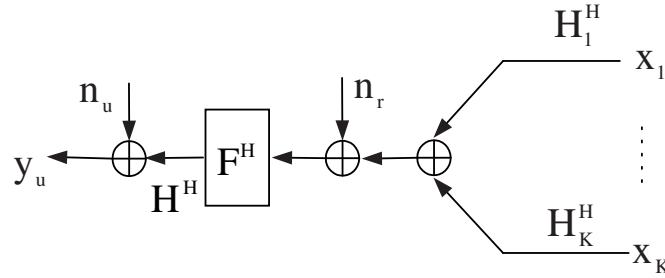


Fig. 2. Diagram of a multiuser MIMO relay uplink system.

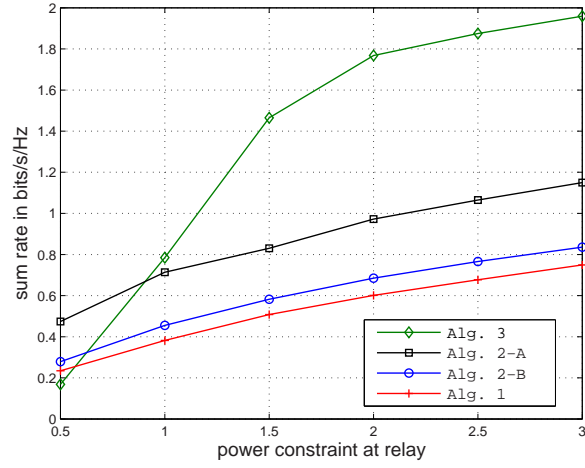


Fig. 3. Comparison of downlink Algorithms 1-3: Averaged sum rate versus power constraint at relay. Alg. 2-A is Algorithm 2 using the best out of 20 random initializations. Alg. 2-B is Algorithm 2 using the results from Algorithm 1 as initializations.

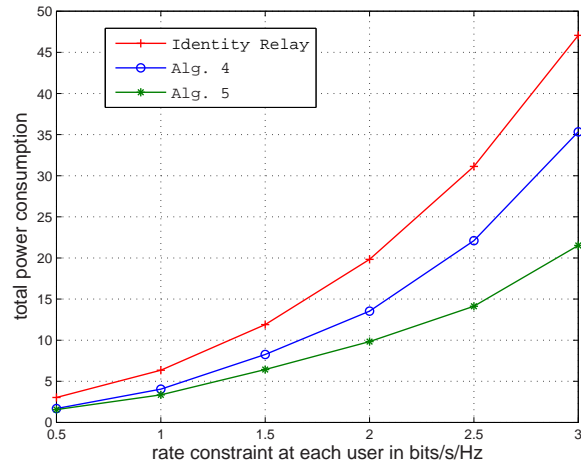


Fig. 4. Comparison of downlink Algorithms 4-5: Averaged total power consumption versus individual rate constraint. The curve on the top is for the identity relay matrix.

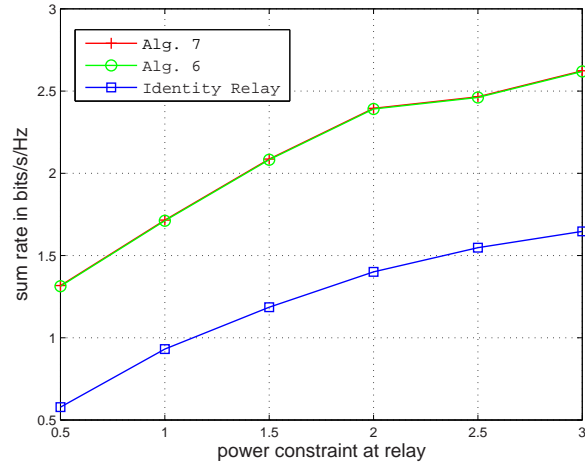


Fig. 5. Comparison of uplink Algorithms 6-7: Averaged sum rate versus relay power constraint. The curves for Algorithms 6-7 are identical. The lower curve is for the identity relay matrix.

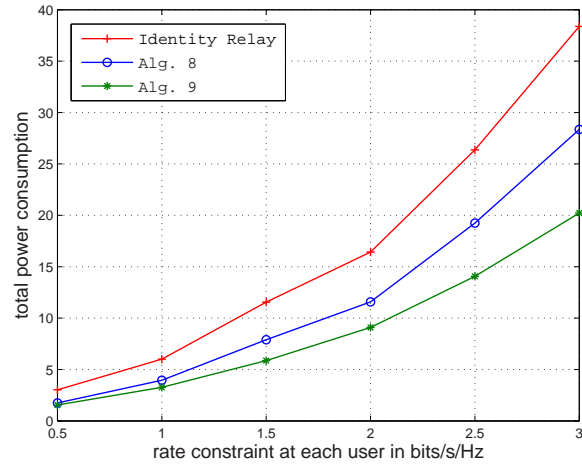


Fig. 6. Comparison of uplink Algorithms 8-9: Averaged total power consumption versus individual rate constraint. The curve on the top is for the identity relay matrix.

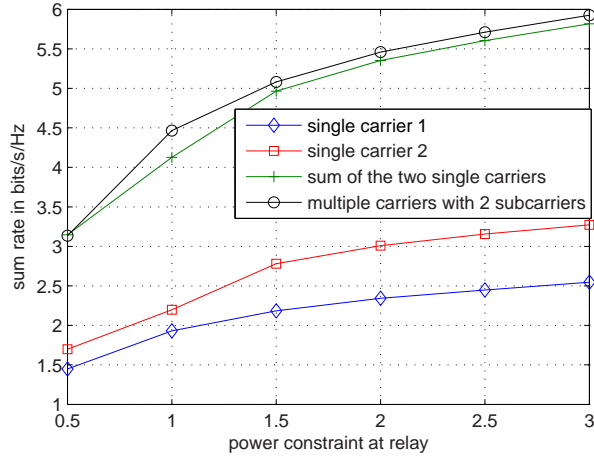


Fig. 7. An example of joint multi-carrier power allocation for downlink multi-user MIMO relay system where  $K = 2$ ,  $N = 2$ ,  $M = 4$  and  $M_c = 2$ . Algorithm 3 was applied with 20 random initializations. The rates shown are based on a single channel realization for each of the two carriers.

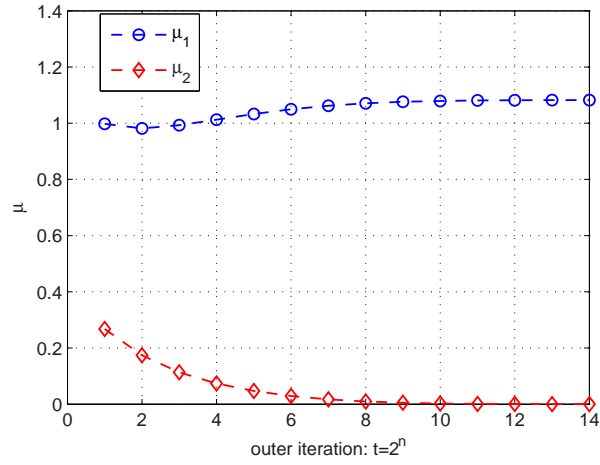


Fig. 8. Optimal values of  $\mu_1$  and  $\mu_2$  as function of the outer loop index  $n$  in  $t = 2^n$ .

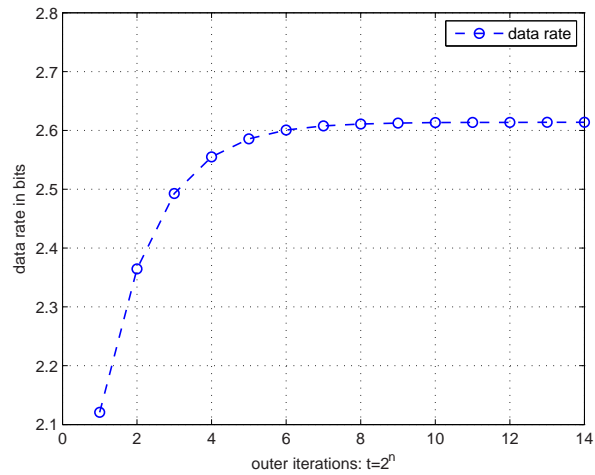


Fig. 9. Optimal value of  $-J$  (capacity) as function of the outer loop index  $n$  in  $t = 2^n$ .

PLACE  
PHOTO  
HERE

**Yuan Yu** received B.E. in Automation from Nankai University, Tianjin, China, in 2002, M.S. in Automatic Control from Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 2005 and Ph.D. degree in electrical engineering from University of California Riverside, Riverside, CA, in 2009.

His research interests include cooperative wireless communications, wireless multi-hop networks, multiuser MIMO communication systems.

PLACE  
PHOTO  
HERE

**Yingbo Hua** (S'86-M'88-SM'92-F'02) received B.S. degree in Control Engineering from Nanjing Institute of Technology (the predecessor of Southeast University), Nanjing, China, in Feb. 1982, and M.S. and Ph.D. degrees in Electrical Engineering from Syracuse University, Syracuse, NY, in 1983 and 1988, respectively.

From 1988 to 1989, he was a research fellow at Syracuse, consulting for Syracuse Research Co., NY, and Aeritalia Co., Italy. He was Lecturer from Feb. 1990 to 1992, Senior Lecturer from 1993 to 1995, and Reader and Associate Professor from 1996 to 2001, with the University of Melbourne, Australia.

He served as a visiting professor with Hong Kong University of Science and Technology from 1999 to 2000, and consulted for Microsoft Research Co., WA, summer 2000. Since Feb. 2001, he has been Professor of Electrical Engineering with the University of California, Riverside, CA.

He is an author/coauthor of numerous articles in journals, conference proceedings and books, which span the fields of sensor array signal processing, channel and system identification, wireless communications and networking, and distributed computations in sensor networks. He is a co-editor of *Signal Processing Advances in Wireless and Mobile Communications*, Prentice-Hall, 2001, and *High-Resolution and Robust Signal Processing*, Marcel Dekker, 2003. He has served on the Editorial Boards for *IEEE Transactions on Signal Processing*, *IEEE Signal Processing Letters*, *IEEE Signal Processing Magazine*, and *Signal Processing (EURASIP)*. He also served on IEEE SPS Technical Committees for Underwater Acoustic Signal Processing, Sensor Array and Multi-channel Signal Processing, and Signal Processing for Communications and Networking, and on numerous international conference organization committees.